

Hit

# ÇEVİRİBİLİMDE DERLEM UYGULAMALARI



Doç. Dr. Elif Tokdemir Demirel



### **Doç. Dr. Elif TOKDEMİR DEMİREL**

Elif Tokdemir Demirel, 1971 yılında Trabzon'da doğmuş, ilk ve orta öğrenimini Trabzon'da tamamlamıştır. Lisans eğitimini 1994 yılında Marmara Üniversitesi İngilizce Öğretmenliği programında, yüksek lisans eğitimini ise 2007 yılında Bilkent Üniversitesi MA-TEFL (Teaching English as a Foreign Language) programında tamamlamıştır. Doktora eğitimini 2009 yılında Orta Doğu Teknik Üniversitesi İngiliz Dili Eğitimi programında tamamlamıştır. 2001–2003 yılları arasında ABD'de Fulbright burslusu olarak Northern Arizona Üniversitesi'nde derlem dilbilimi alanında araştırmalar yürütmüştür. Hâlen Kırıkkale Üniversitesi Batı Dilleri ve Edebiyatları Bölümü İngilizce Mütercim ve Tercümanlık Anabilim Dalı Başkanı olarak görev yapmaktadır. Akademik çalışmaları derlem dilbilimi, çeviri çalışmaları, İngiliz dili eğitimi, akademik yazma, uygulamalı dilbilim ve dijital beşerî bilimler alanlarında yoğunlaşmakta; uzun yıllardır lisans ve lisansüstü düzeyde İngiliz dili eğitimi ve çeviribilim alanlarında dersler vermektedir

Kırıkkale Üniversitesi, Batı Dilleri ve Edebiyatları Bölümü

İngilizce Mütercim ve Tercümanlık Anabilim Dalı

Yahşihan, Kırıkkale

Tel: 0531 7931922

e-posta: e.demirel@kku.edu.tr, [elif6171@gmail.com](mailto:elif6171@gmail.com)

ORCID ID: 0000-0003-1280-2679

# ÇEVİRİBİLİMDE DERLEM UYGULAMALARI

Doç. Dr. Elif TOKDEMİR DEMİREL

**ÇİZGİ**  
KİTAP E Vİ

Çizgi Kitabevi Yayınları (e-kitap)

©Çizgi Kitabevi  
Aralık 2025




ISBN: 978-625-396-759-8  
Yayıncı Sertifika No: 52493

**KÜTÜPHANE BİLGİ KARTI**  
- Cataloging in Publication Data (CIP) -  
TOKDEMİR DEMİREL, Elif  
ÇEVİRİBİLİMDE DERLEM UYGULAMALARI

## ÇİZGİ KİTABEVİ

Sahibiata Mah. M. Muzaffer Cad. No:41/1 Meram/ <b>Konya</b> (0332) 353 62 65	Konevi Mh. Larende Cad. No:20/A Meram/ <b>Konya</b> (0332) 353 62 66	Siyavuşpaşa Mh. Gül Sk. No: 15 B Bahçelievler/ <b>İstanbul</b> (0212) 514 82 93
---	---	--

[www.cizgikitabevi.com](http://www.cizgikitabevi.com)

   / cizgikitabevi

# İÇİNDEKİLER

ÖNSÖZ.....	7
------------	---

---

## *BÖLÜM I – DERLEM DİLBİLİMİNE GİRİŞ*

---

1.1. Tarihçe ve Gelişim .....	10
1.2. Temel Kavramlar ve Terimler .....	12
1.3. Derlem Türleri .....	21
1.3.1. Tek Dilli, İki Dilli ve Çok Dilli Derlemler .....	22
1.3.2. Paralel Derlemler .....	24
1.3.3. Karşılaştırılabilir Derlemler .....	28
1.3.4. Özel Amaçlı Derlemler .....	29
a. Sözcükbilim Derlemleri .....	29
b. Sözlük Derlemler .....	30
c. Sözlü İngilizce Derlemleri .....	32

---

## *BÖLÜM II – DERLEM DİLBİLİMİ VE ÇEVİRİBİLİM*

---

2.1. Betimleyici Çeviri Çalışmaları ve Derlem Dilbilimi.....	37
2.2. Veriye Dayalı Çeviribilim.....	43

---

## *BÖLÜM III - ÇEVİRİBİLİM ARAŞTIRMALARI VE UYGULAMALARINDA DERLEM KULLANIMI*

---

3.1. Çeviribilim Odaklı Derlem Tasarımı .....	47
3.2. Planlama Adımları .....	47
a. Yazılım Seçimi .....	48
b. Derlem Büyüklüğü .....	50
c. Metin Seçimi ve Matris Planlaması.....	51
d. Paralel Derlemlerde Temsil Edilebilirlik.....	51
e. Metin Girişlerinin Yapılması .....	52
f. Metinlerin Hizalanması.....	52

g. Metin Anotasyonu .....	54
h. Derlemin Depolanması .....	64
i. Telif Hakkı Sorunları .....	65
3.3. Çeviribilimde Derlem Analizi için Araçlar ve Teknikler .....	66
3.3.1. Derlem Üzerinde Arama Yapmak .....	67
3.1.2. KWIC Formatı (Bağlamda Anahtar Kelime Gösterimi) .....	75
3.1.3. Cümle Tabanlı Bağlam Dizinleri .....	76
3.1.4. Paralel Bağlam Dizinleri .....	77
3.1.5. Yazılım Araçları .....	78

---

#### *BÖLÜM IV. ÇEVİRİBİLİM ARAŞTIRMALARINDA DERLEM KULLANIMI*

---

4.1. Derlem Verilerinde Merkezi Eğilimlerin Araştırılması .....	81
4.2. Derlem Tabanlı Araştırmada Anahtar Kelime Analizi .....	82
4.3. Derlem Tabanlı Araştırmada Eşdizimlilik Analizi .....	84
4.4. Sözcükbilim Analizlerinde Paralel Derlem Kullanımı .....	86
4.5. Çokdilli derlemler ve Terminoloji Çalışmaları .....	88
4.6. Paralel Derlem Araştırmalarında Morfoloji ve Sözdizimi .....	90
4.7. Edimbilim ve Çok Dilli Derlem Analizi .....	91

---

SONUÇ .....	94
-------------	----

---

KAYNAKÇA .....	96
----------------	----

---

#### *EKLER*

---

• Ek A- Anahtar Terimler Sözlüğü .....	105
• Ek B- Alıştırmalar .....	109
• Ek C- Biber etiketleyici tarafından atanan etiketlerin listesi .....	112

---

# Önsöz

Son yıllarda, dijital dil verilerinin ve hesaplama araçlarının artan erişilebilirliği, dil ve çeviri üzerine yapılan araştırmaların biçimini kökten etkilemiştir. Önceleri büyük ölçüde sezgi odaklı olan bu alan, zamanla daha ampirik, veri odaklı ve metodolojik olarak şeffaf hale geldi. Bu değişim, özellikle çeviri çalışmaları için önemli olmuştur, burada derlem tabanlı yaklaşımlar çeviriyi tanımlamak, açıklamak ve öğretmek için yeni yollar açmıştır.

Bu kitabın temel motivasyonu, çeviri araştırması ve uygulamasıyla ilgilenen okuyucular için derlem dilbilimine yapılandırılmış ve erişilebilir bir giriş sunmaktır. Derlem dilbilimi ve derlem tabanlı çeviri çalışmaları üzerine önemli bir literatür bulunmaktadır, ancak bu eserler genellikle alanlar arasında dağınık ve farklı seviyelerde bilgiler sunmaktadır. Bu kitap, temel kavramları, metodolojik ilkeleri ve Pratik uygulamaları tutarlı ve bütünlük bir şekilde sunarak bu yönleri bir araya getirmeyi amaçlamaktadır. Bu kitap, çeviri çalışmaları, dilbilim ve ilgili alanlarda lisansüstü öğrenciler, araştırmacılar ve uygulayıcılar için oluşturulmuştur.

Kitap dört ana bölüme ayrılmıştır. İlk bölüm derlem dilbilimini tanıtır, tarihsel gelişimini, temel kavramlarını ve başlıca derlem türlerini özetler. Temsil edilebilirlik ve derlem tasarımı gibi temel konular, sonraki bölümler için sağlam bir temel oluşturmak amacıyla tartışılmaktadır. İkinci bölüm, derlem dilbilimi ile çeviri çalışmaları arasındaki ilişkiyi inceleyerek, derlem tabanlı yaklaşımları betimleyici ve veri odaklı çeviri araştırmaları içinde konumlandırır. Özellikle, çeviri dil,

normlar ve çeviri evrenselleri etrafındaki tartışmalara odaklanır.

Üçüncü bölüm, çeviri araştırmasında derlem kullanımının pratik yönlerine odaklanır. Okuyucuya planlama, metin seçimi, hizalama, açıklama, depolama ve telif hakkı konuları dahil olmak üzere çeviri odaklı derlemlerin tasarım ve derlenme aşamalarında rehberlik etmektedir. Yöntemsel yetkinlik ve araştırma özerkliğini teşvik etmek amacıyla yaygın kullanılan araçlar ve analitik teknikler de tanıtılmaktadır. Dördüncü bölüm, derlem verilerinin çeviri çalışmalarında sözcüksel, dilbilgisel, pragmatik ve çok dilli analizler dahil somut araştırma sorularına nasıl uygulanabileceğini gösterir.

Bu kitabın hem derlem tabanlı çeviri çalışmalarına yeni başlayanlar için bir giriş hem de daha ileri araştırmalar için pratik bir referans olarak hizmet etmesi ümit edilmektedir. Sonuç olarak amaç karmaşık dilsel ve sosyal bir uygulama olan çevirinin daha derin anlaşılmasına katkıda bulunan titiz, şeffaf ve anlamlı ampirik çalışmaların geliştirilmesini desteklemektir.

Doç. Dr. Elif Tokdemir Demirel

*Bana çalışma azmi ve yaşam sevinci aşıl原因an sevgili babam  
Öğretmen Mikail Tokdemir'in anısına.*

# BÖLÜM I - DERLEM DİLBİLİMİNE GİRİŞ

*“You shall know a word by the company it keeps.”*

J. R. Firth

## 1.1. Tarihçe ve Gelişim

Dil arařtırmaları zaman içinde yapısal arařtırmalar ve kullanım odaklı arařtırmalar olmak üzere iki temel alana odaklanmıřtır. Geleneksel olarak, dilbilim arařtırmaları yapısal alana odaklanmıřtır ve dilin yapısal birimlerini ve yapısal sınıflandırmalarını tanımlamayı hedeflemiřtir. Yapısal arařtırmaların amacı biçimbirim, kelime, sözdizimsel öbek ve dilbilgisel sınıflandırmaları tanımlamak ve bu birimlerin daha büyük dilbilgisel ögeler oluşturmak için nasıl bir araya gelebileceklerini betimlemektir. Örneğın,

- *Kelimeler birleşerek nasıl sözdizimsel öbekler haline gelmektedir? veya*
- *Sözdizimsel öbekler birleşerek nasıl tümcecikler haline gelmektedir?*

Diğeri yaklaşım olan kullanım odaklı yaklaşım ise dilin doğal ortamında kullanımına odaklanır. Bu yaklaşımda, herhangi bir dilin konuşucularının o dilin kaynaklarını nasıl kullandıkları ve herhangi bir dilde teorik açıdan nelerin mümkün olduđu sorusundan çok dilin doğal olarak oluşturulmuş sözlü veya yazılı metinlerde gerçek anlamda nasıl kullanıldıđı arařtırma konusudur (Biber vd., 2000). Bu bağlamda, kullanım odaklı arařtırmalar dil kullanımı hakkında farklı sorular sorabilirler. Örneğın,

- *“Bir dilde neden anlam ve dilbilgisel işlevi aynı olan farklı yapılar vardır?”*

• *“Konuşma dili ve yazı dilinde farklı dil yapıları bilinçli olarak mı tercih edilmektedir?”*

yoksa

• *“Farklı yapılar özelleşmiş anlamları ifade etmek için mi tercih edilmektedir?”*

Bu tür yapısal soruların yanında kullanımsal farklar da ilgi çekici olabilir. Örneğin,

• *“Belirli bir yazarın dil kullanımı onun çağdaşları ile karşılaştırıldığında nasıl bir değişkenlik gösterir?”*

• *“Kadınlar tarafından kullanılan dil ile erkekler tarafından kullanılan dil arasında nasıl farklar vardır?”*

Bunlara ek olarak farklı metin türleri ve dil dizgeleri arasındaki farklılaşmayı ilgilendiren sorular da sorulabilir. Örneğin,

• *“Karşılıklı konuşma, gazete, akademik makale, kişisel mektup gibi dil dizgelerinde kullanılan farklı dil kalıplarını nasıl ayırt edebiliriz?”*

Bütün bu kullanım sorularına cevap bulmak amacıyla yapılan araştırmalarda iki temel araştırma hedefi vardır:

• *Herhangi bir kullanım kalıbının ne kadar sıklıkla kullanıldığı ve*

• *Bu kalıpların değişkenliğini etkileyen bağlamsal faktörlerin neler olabileceği.*

Kullanım kalıplarını bulmak için yapılan araştırmadaki başlıca yöntemsel zorluk, tipik olarak tercih edilen kalıplar bulunmak istendiğinden analizlerin sistematik olmayan kişisel tecrübelerle dayalı verilerle veya sezgilere dayanarak yapılamamasıdır. Dil incelendiğinde çoğunlukla tipik kullanımlardan çok olağan olmayan kullanımlar göze çarpar ve bu sebeple de sezgisel çıkarımlar güvenilir olmaz. Güvenilir

çıkarımlar yapılabilmesi ve çıkarımların sadece kişiye özgü farklılıklara dayandırılmaması için birçok farklı dil konuşucusundan toplanmış çok sayıda dil örneklerine ihtiyaç olacaktır. Çok büyük ölçekte dil örneklerinin ise klasik yöntemlerle işlenebilmesi çok zaman alıcı ve zahmetli bir süreç olacaktır. Bu zorluklar dil araştırmacılarını hem daha güvenilir hem de dil verisini daha hızlı işleyebilmeye olanak sağlayacak yöntemler bulmaya yönlendirmiştir. Bu bağlamda geliştirilen derlem-tabanlı yaklaşım hem büyük miktarda dil verisini güvenilir şekilde ve hızla analiz etmeyi hem de aynı anda birçok farklı bağlamsal faktörü kontrol edebilmeyi sağlamıştır.

## 1.2. Temel Kavramlar ve Terimler

Derlem teriminin İngilizcesi olan 'corpus' kelimesi Latince'de 'gövde, vücut' anlamlarına gelir. Derlem sözlü ve yazılı metinlerden oluşan, belirli ilkelere göre düzenlenmiş, farklı metin türlerini içeren ve söylem unsurlarına göre kodlanmış bir metinler topluluğu olarak tanımlanabilir (Biber vd., 2006). Kennedy (2014) bir derlemi yazılı veya sözlü belirli metin örneklerini temsil eden, dilbilimsel araştırma da dahil olmak üzere çeşitli amaçlar için derlenmiş, prozodi, dilbilgisi, söylem kalıpları gibi çeşitli dil seviyelerindeki soruları yanıtlamak için tasarlanmış pragmatik bir metin koleksiyonu olarak tanımlamıştır. İlk derlemlerin elektronik olmamalarına rağmen günümüzde bilgisayar tarafından okunabilen derlemler ön plandadır ve derlem elektronik bir metin koleksiyonu olarak kabul görmektedir.



Şekil 1. Derlem Dilbilimi Zaman Çizelgesi

Bilgisayar aracılığıyla derlem oluşturulması ilk olarak 1960'lı yıllarda Brown derlemi ile başlamıştır (Meyer, 2002). Brown Derlemi (resmi adıyla Brown Üniversitesi Günümüz Amerikan İngilizcesi Standart Derlemi), modern İngilizce üzerine dilbilimsel araştırmalar için bilgisayarda okunabilir ilk genel derlemdir. 1960'ların başında Brown Üniversitesi'nde W. Nelson Francis ve Henry Kuçera tarafından derlenen bu eser, derlem dilbilimi için temel bir kaynak olmaya devam etmektedir.

Derlemin boyut olarak yaklaşık 1.014.312 kelimededen oluşmakta ve her biri yaklaşık 2.000+ kelimededen oluşan 500 metin örneği içermektedir. Brown Derlemi Amerika Birleşik Devletleri'nde 1961'de yayımlandı ve başlangıçta delikli kartlar şeklinde depolanan derlem, artık konuşma bölümleriyle etiketlenmiş versiyonlar da dahil olmak üzere birden fazla dijital formatta mevcuttur (VARIENG, t.y.). O zamandan beri, çeşitli boyutlarda ve çeşitli amaçlarla çok sayıda derlem oluşturulmuş ve derlemlerin analizine yardımcı olan çok sayıda bilgisayar programı geliştirilmiştir. Bazı derlem örnekleri aşağıdaki Tablo 1'de verilmiştir.

Tablo 1: Derlem Kaynakları (Meyer, 2002)

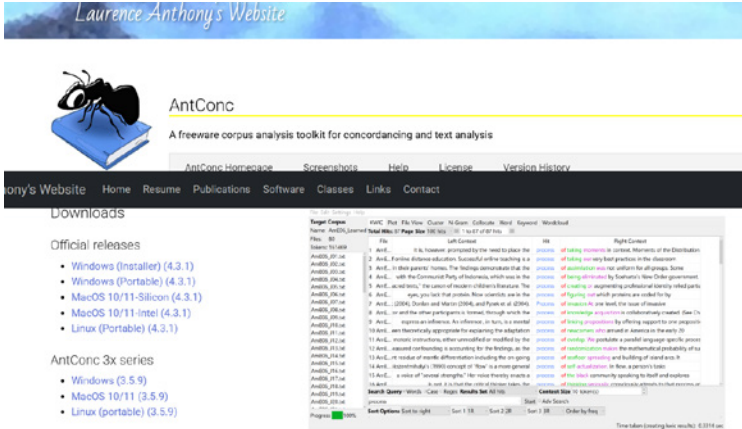
Kaynak	Açıklama	Erişim
<b>Amerikan Ulusal Derlemi</b>	Devam ediyor: sözlü ve yazılı metin örnekleri.	<b>Proje internet sitesi:</b> <a href="http://www.cs.vassar.edu/~ide/anc/">http://www.cs.vassar.edu/~ide/anc/</a>
<b>İngilizce Bankası Derlemi</b>	415 milyon kelime konuşma ve yazma (Ekim 2000 itibariyle); metinler sürekli olarak eklenmektedir.	<b>Collins-Cobuild:</b> <a href="http://titania.cobuild.collins.co.uk/boe_info.html">http://titania.cobuild.collins.co.uk/boe_info.html</a>
<b>Birmingham Derlemi</b>	20 milyon kelime yazılı İngilizce örnekleri.	<b>Bank of English Derlemine dönüştü</b>
<b>İngiliz Ulusal Derlemi</b>	Sözlü (10 milyon) ve yazılı (90 milyon kelime) İngiliz İngilizcesi içeren farklı uzunluklarda 100 milyon kelimelik örnek.	<b>BNC internet sitesi:</b> <a href="http://info.ox.ac.uk/bnc/index.html">http://info.ox.ac.uk/bnc/index.html</a>
<b>Brown Derlemi</b>	Bir milyon kelimelik düzenlenmiş Amerikan İngilizcesi örnekleri.	<b>ICAME CD-ROM</b>
<b>Cambridge Uluslararası Derlemi</b>	100 milyon kelime sözlü ve yazılı İngiliz ve Amerikan İngilizcesi örnekleri.	<b>CUP İnternet sitesi:</b> <a href="http://uk.cambridge.org/bnc/getting/sampler.html">http://uk.cambridge.org/bnc/getting/sampler.html</a>
<b>Cambridge Öğrenci Derlemi</b>	10 milyon kelimelik öğrenci kompozisyon sınav metin örnekleri.	<b>CUP İnternet sitesi:</b> <a href="http://uk.cambridge.org/bnc/getting/sampler.html">http://uk.cambridge.org/bnc/getting/sampler.html</a>
<b>Hong Kong Bilim ve Teknoloji Üniversitesi Öğrenci Derlemi</b>	25 milyon kelime İngilizce öğrenen öğrencilerin dil örnekleri.	<b>İletişim: John Milton Project Director</b> <a href="mailto:lsjohn@ust.hk">lsjohn@ust.hk</a>
<b>Uluslararası İngilizce Derlemi (ICE)</b>	600.000 kelimelik konuşma, ulusal İngilizce çeşitlerini temsil eden 400.000 kelimelik metin örnekleri.	<b>Great Britain:</b> <a href="http://www.ud.ac.uk/english-us-age/ice-gb/index.htm">http://www.ud.ac.uk/english-us-age/ice-gb/index.htm</a>

Derlem dilbilim yaklaşımı ile dil analizine yönelik üretici-dönüşümsel dilbilgisi (Chomsky, 1968) gibi geleneksel yaklaşımlar arasındaki temel fark, 'kuralcılık' ve 'betimleme' arasındaki ayrımdır (Meyer, 2002). Derlem dilbilimcileri, günlük kullanımdan elde edilen verilere dayanarak dilin gerçek kullanımını tanımlamaya çalışırken, üretici-dönüşümsel

dilbilimciler ideal dil teorileri geliştirmeyi ve böylece dilin doğru veya yanlış kullanımının ne olduğunu belirlemeyi amaçlar. Meyer'den (2002) yapılan aşağıdaki alıntı bu farkı daha da açıklamaktadır:

**Derlem dilbilimcisi için birincil kaygı, dilin doğru bir tanımıdır; üretici-dönüşümsel dilbilgisi için önemli olan, evrensel dilbilgisi bilgilerimizi geliştiren teorik bir dil tartışmasıdır. (Meyer, 2002, s, 4)**

Derlem dilbilimi, toplumsal bir varlık olan dili, bilimsel olarak analiz edilebilir bir şekilde getirerek dilbilim araştırmalarına büyük katkı sağlamıştır. Dil analizi için özel olarak tasarlanmış Monoconc (Thomas & Short, 1994), Sara (Burnard, 2000), AntConc (Anthony, 2019) gibi bilgisayar programları aracılığıyla günlük dilin büyük çevrimiçi örneklerini analiz etmek ve belirli dil öğelerinin bir arada ortaya çıkması gibi dilin çeşitli özelliklerini araştırmak mümkün hale gelmiştir. Başlıca avantajı, dilbilimsel analizlerin derlem dilbilimi aracılığıyla her zamankinden daha büyük bir ölçekte yapılabilmesidir



Şekil 2: AntConc yazılımı (Anthony, 2019)

Bir derlem kullanılarak yazılı veya sözlü dil üzerinde ampirik arařtırmalar yapılabilir. Geleneksel dil arařtırmalarında tek metinler edebi veya dilbilimsel arařtırmalara temel oluşturabilir. Fakat, derlem arařtırmaları tek metne dayalı arařtırmalardan farklıdır. Tek bir metinden fazla sayıda metinden oluşan dil örnekleri koleksiyonlarına derlem denilebilse de modern dilbilimde derlem kelimesinin çağrıřtırdığı farklı anlamlar vardır. Tam anlamıyla bir dil koleksiyonuna derlem denilebilmesi için bu koleksiyonun birtakım özelliklere sahip olması gerekir. Bunlar temelde örnekleme, temsil edebilirlik, sonlu sayıda örnek içermeye, bilgisayarca okunur olma ve standart bir başvuru kaynağı oluşturabilme özellikleridir.

Dilbilimciler dili tek metinler üzerinden değil dilin farklı varyasyonları üzerinden arařtırlar. Burada karşımıza izlenebilecek iki yol çıkar: bunlardan birincisi her bir tümceyi analiz etmektir; fakat bu yaşayan bir dili arařtırmada izlenebilecek mantıklı bir yol değildir. Bunun yerine ikinci yol olan bir dil varyasyonunun bir örneklemini oluşturmak seçilebilir. Bu bağlamda günümüzdeki kadar geniş temsiliyet gücüne sahip olmayan ilk derlemler Chomsky gibi geleneksel dilbilimciler tarafından yanılıcı oldukları şekline eleřtirilmiřtir. Günümüzde ise modern derlemler bu eleřtiriye bořa çıkaracak şekilde daha büyük, daha kapsamlı ve daha temsiliyet gücü yüksek hale gelmiřlerdir. Burada amaç dilde var olan eğilimlerin doğru bir temsilini yapabilmektir. Örneğin 19. yy. İngilizcesini betimlemek isteyen bir derleme sadece Dickens veya Bronte romanlarını dahil etmek yetersiz kalacaktır. Daha dengeli bir temsiliyet elde edebilmek için bir derlem birden fazla yazar ve metin türü içermelidir.

Bir derlem sonlu sayıda dil örneği içerebilir; örneğin birçok derlem bir milyon kelime gibi önceden belirlenmiř bir büyüklüğe sahiptir. Monitör derlemler olarak adlandırılan bazı

derlemler ise belirli bir büyüklüğe sahip değil ucu açık olarak tasarlanmışlardır. Bu tür derlemlere örnek olarak Birmingham Üniversitesi'ndeki COBUILD derlemi (Groom, 2013) gösterilebilir. Bu sözdizimsel açıklamalı dil kaynağı, Helsinki Üniversitesi ile Alabama Üniversitesi Birmingham (UAB) arasında iş birliği yoluyla geliştirilmiştir. Hem POS etiketli metin derlemi hem de yazılı ve sözlü modaliteleri kapsayan bir konuşma derleminden oluşur. Veriler tek dillidir ve tamamen İngilizcedir; morfoloji ve sözdizimin dilsel alanlarına odaklanmaktadır. Kaynak, Glasgow'daki HarperCollins Publishers tarafından desteklenen ve Birmingham Üniversitesi'ndeki COBUILD ekibi tarafından yürütülen uluslararası bir İngilizce dil projesi olan Bank of English'e dayanmaktadır. Derlem, yaklaşık 200 milyon yazılı ve sözlü İngilizce kelimedenden oluşmaktadır. 1993–1994 yılları arasında, tüm veriler, Helsinki Üniversitesi Hesaplamalı Dilbilim Araştırma Birimi'nde (RUCL) İngilizce morfolojik analizör ENGTWOL ve İngilizce Kısıtlama Dilbilgisi ayrıştırıcısı ENGCG kullanılarak morfolojik ve sözdizimi olarak etiketlenmiştir. Açıklama süreci, ardışık ön işleme, sözcük analizi, morfolojik anlam ayrımı ve sözdizimi haritalama aşamalarını içeriyordu. İlk 103 milyon kelime 1993'te işlendi. Proje, Birmingham'da Profesör John Sinclair ve Helsinki'de Profesör Fred Karlsson tarafından yönetildi; açıklamayı Timo Järvinen yürüttü. Devam eden analizler, gelen verilerin ve açıklamalı çıktılarının düzenli izlenmesini, ayrıca ön işleme modülü ve ENGTWOL sözlüğünün güncellemeleriyle yakından bağlantılı dokümantasyon ve veri yönetimi görevlerini içeriyordu. Tam kaynak çevrimiçi olarak mevcuttur ve daha fazla bilgi proje iletişim sayfasından erişilebilir (Järvinen, t.y.).

Bir monitör derleminin sonlu derlemlere göre bazı avantajları vardır; örneğin yeni metinler ekleyebilme, dil hakkında

geniş kapsama alanına sahip olma gibi. Diğer taraftan monitör derlemler nicel arařtırmalarda sonlu derlemler kadar güvenilir olmayabilir ve derlenme süreçleri daha rastgele olabilir. Monitör derlemlerin aksine sonlu derlemlerde veri toplama belli bir sayıya ulařıldığında sonlandırılır. Örnek olarak Brown Derlemi (1 milyon kelime) (Francis ve Kućera, 1982), LOB Derlemi (1milyon kelime) (Johansson, Leech ve Goodluck, 1978) ve BNC Derlemi (Burnard, 2000).

Geçmiřte derlem kelimesi basılı metinler anlamında kullanılırken, günümüzde neredeyse bütün derlemler bilgisayarca okunur yani dijital durumdadır. Derlemlerin bilgisayarca okunabilir olması derlem içinde arama yapmada ve anotasyon eklemeye hız ve kolaylık sağlar. Standard derlemler genellikle bir dil varyasyonu için bir başvuru kaynağı olarak kabul edilirler. Bu yaklaşımın da çalışmalar arasında karşılařtırmaya olanak sağlaması ve dilbilimsel arařtırmalar için bir veri bankası sağlaması gibi avantajları vardır. Brown Derlemi (Amerikan İngilizcesi), LOB Derlemi (İngiliz İngilizcesi) ve London-Lund Derlemi (Sözlü İngiliz İngilizcesi) bu tür derlemlere örnek olarak gösterilebilir (Francis ve Kućera, 1964; Johansson vd., 1978; Svartvik, 1990). London-Lund Derlemi (LLC), University College London'daki İngilizce Kullanımı Arařtırması ile Lund Üniversitesi'ndeki Konuşma İngilizce Arařtırması arasında iş birliğı sonucu geliştirilen iyi bilinen bir Britanya İngilizcesi derlemidir. Derlem, ağırlıklı olarak 1950'ler ile 1980'ler arasında kaydedilmiş yaklaşık 500.000 kelimelik sözlü veriden oluşmaktadır. Yüz yüze konuşmalar, telefon görüşmeleri ve hazırlıklı konuşma gibi geniş bir konuşma yelpazesini içerir. London-Lund Derlemi'nin en ayırt edici özelliklerinden biri, duraklama, vurgu, tonlama ve diğer konuşma dili özelliklerini işaretleyen ayrıntılı prosodik transkripsiyonudur. Bu da derlemi özellikle konuşma dili

üzerine söylem, pragmatik ve tonlama kalıpları üzerine arařtırmalar için deęerli kılar. En erken sistematik olarak derlenen konuřma derlemlerinden biri olan London–Lund Derlemi, konuřma dilinin derlem tabanlı alıřmalarının geliřtirilmesinde öncü bir rol oynamıřtır (Svartvik, 1990).

Bir derlem sadece rastgele bir metin koleksiyonu deęildir, sistemli örnekleme ieren, temsiliyet gücü olan, bilgisayarca okunabilir ve standart bařvuru saęlayan sistemli bir dil koleksiyonudur. Farklı derlemler oluřturulma amalarına baęlı olarak farklı arařtırma amaları için kullanılabilirler. Derlem temelli analizler sadece dilsel özelliklerin basit bir sayımı deęildir. Derlem tabanlı bir yaklařım, arařtırmacıların birçok analiz yapmasına olanak tanır. Bir derlem kullanılarak bir dil kapsamında öęeler arasındaki sözcüksel ve dilbilgisel karmařık iliřki kalıplarını analiz edilebilir. Farklı dil dizgeleri, leheler, zaman dilimleri arasında daęılım üzerine karřılařtırma analizler yapılabilir, sözlükler gibi referans alıřmaları geliřtirilebilir (örn. Cobuild Sözlüęü), dil öęretim materyalleri geliřtirilebilir, dil edinimi üzerine arařtırmalar yapılabilir veya eviribilim için karřılařtırmalı diller arası analizler yapılabilir (Kennedy, 2014).

Bilgisayarın da kullanımı ile derlem temelli yaklařım, arařtırmacılara dilde bulunan karmařık iliřki kalıplarını, yani dilbilimsel unsurların birbirleri ile ve dięer dil dıřı unsurlarla olan sistematik iliřkilerini belirleme ve analiz etme fırsatı sunan bir arařtırma yöntemidir.

Biber vd. (2006) derlem dilbilimi yaklaşımının özelliklerini aşağıdaki şekilde sıralamıştır:

3. Bilgisayarlardan yoğun biçimde yararlanır; analiz için hem otomatik hem de etkileşimli teknikleri kullanır.

4. Hem nicel (quantitative) hem de nitel (qualitative) çözümlene tekniklerine dayanır.

Bu özellikler bir araya geldiğinde, derlem temelli yaklaşım başka şekilde mümkün olmayan kapsamlı ve güvenilir bir analiz sağlar. Derlem temelli yaklaşımın avantajlarından bazıları bilgisayar kullanımından kaynaklanır. Bilgisayarlar, karmaşık örüntüleri tanımlamayı ve analiz etmeyi mümkün kılar, daha büyük bir doğal dil örneğinin depolanmasına ve analizine olanak tanır; bu da elle yapılabilecek olandan fazladır. Bilgisayarlar tutarlı ve güvenilir analizler sunar – insanlar analiz sırasında zihin olarak yorulabilir ya da sıkılabilir. Bilgisayarlar ise etkileşimli olarak kullanılabilir, bu sayede insan analizcinin dilsel yargılarda bulunmasına olanak tanırken, bilgisayar işlemi yürütür. Ayrıca, derlem temelli yaklaşım kullanılan dil analizlerinde bilgisayar üzerindeki çevrimiçi derlem verilerinin her yerinden araştırmacılara erişim sağlar ve derlem verileri dil öğrenenler için ideal ve gerçekçi veriler sağlar (Startvik,1992).

Svartvik (1992), dil analizinde Derlem temelli bir yaklaşımın avantajlarını aşağıdaki şekilde sıralamıştır:

#### Derlem Verileri:

1. Objektiftir.
2. Kolayca doğrulanabilir veriler içerir.
3. Lehçeler, dil dizgeleri ve stiller arasındaki değişkenlik çatışmaları yapılmasına olanak sağlar.
4. Dilsel öğelerin oluşma sıklığını araştırmaya olanak sağlar
5. Dilsel öğelerin oluşma sıklığını araştırmaya olanak sağlar.
6. Açıklayıcı örnekler sunar.
- 7. Dil öğretimi, makine çevirisi, konuşma sentezi vb. gibi bir dizi uygulamalı alanı için gerekli veriler sağlar

### 1.3. Derlem Türleri

Dilbilimsel analizler yapmak amacıyla oluşturulmuş birçok farklı metin içeren elektronik derlemler mevcuttur. Bu derlemler oluşturulma amaçlarına, temsil ettikleri metin türüne, yapılarına ve formatlarına göre farklılık gösterirler. Bazı derlemler sadece farklı konularda dilbilim çalışmaları için bir metin altyapısı sağlamak amacıyla derlenmişlerdir. Bu tür derlemler '*genel derlemler*' olarak adlandırılabilir ve dilbilimciler tarafından dilin kelime, gramer veya söylem yapısı gibi alanlarda öne attıkları sorulara cevap aramak için analiz edilebilecekleri metinlerden oluşurlar. Bu sorular örneğin İngilizcenin farklı varyasyonları olan Amerikan İngilizcesi ve İngiliz İngilizcesi arasındaki farklarla ilgili sorular olabilir. Belirli araştırma projelerini yürütmek için derlenmiş olan derlemler ise '*özelleşmiş (specialized) derlem*' olarak adlandırılabilir. Örneğin, büyük yayıncılar tarafından kelime sıklıkları hakkında bilgi edinmek veya modern sözlükler oluşturmak için atfı sağlamak amacıyla oluşturulan derlemler bu türe dahil edilebilir.

Leech (1992) dil modelleri oluşturulmasını ve dil işlemeyi kolaylaştırmak amacıyla derlenen özelleşmiş derlemleri '*eğitici derlem*' veya '*test derlemi*' olarak adlandırmıştır. Özelleşmiş derlemlerin başlıca türleri arasında bölgesel veya sosyodilbilimsel farkları incelemek için derlenmiş olan derlemler sayılabilir. Bu kategorinin altında 'lehçe derlemleri', 'bölgesel derlemler', 'öğrenci derlemleri' veya 'standard-dışı derlemler' eklenebilir. Diğer taraftan konuşma dili dil kullanımında en yaygın mod olmasına rağmen derlem tabanlı gramer ve kelime çalışmalarının büyük çoğunluğu yazılı derlemlerin analizine dayanmaktadır. Derlem dilbilimi literatüründe, birkaç farklı derlem türü öne çıkmaktadır. Bunlar arasında **tek dilli, iki dilli, çok dilli derlemler, karşılaştırılabilir derlemler ve paralel derlemler** sayılabilir.

### 1.3.1. Tek Dilli, İki Dilli ve Çok Dilli Derlemler

Birçok derlem tek bir dilde metinlerin veya aynı dilin farklı varyasyonlarının derlenmesi ile oluşturulmuştur, örneğin, Uluslararası İngilizce derlemi (ICE: International Corpus of English) (Greenbaum, 1996). Paralel metin derlemleri söz konusu olduğunda ise en yaygın olan tür sadece iki dilde metinler içeren derlem türüdür. Fakat birkaç farklı dilde metinler içeren paralel metin derlemleri de mevcuttur. Özellikle, farklı birkaç dilde eşleşen metinler bulmanın zorluğundan dolayı çok sayıda dilde metinler içeren paralel derlemler oluşturmak hem zaman alıcı hem de masraflı bir iştir. Bu sebeple, kaçınılmaz olarak çok dilli derlemler iki dilli derlemlere kıyasla daha küçük ve daha dar kapsamlı kalmaktadır. Fakat küçük veya dar kapsamlı olmalarına rağmen, belirli alanlardaki araştırmalarda (örn. Dil tipolojisi alanında) çok faydalı bilgiler sağlayabilmektedirler.

Çok Dilli derlemler orijinal dillerinde yazılmış metinlerden ve/veya farklı dillerden çevirilerden oluşurlar. Bu derlemler farklı çeşitlemeler ile karşımıza çıkar. Bu derlemler aşağıda da görüldüğü gibi Teubert (1996) tarafından şu şekilde isimlendirilmiştir: (a) karşılaştırma derlemi , (b) paralel derlem, (c) çeviri derlemi.

(a) A dilinde orijinal metinler ve B dilinde (farklı) orijinal metinler. (karşılaştırma derlemi)

(b) A dilinde original metinler ve bu metinlerin B diline çevirileri. (paralel derlem)

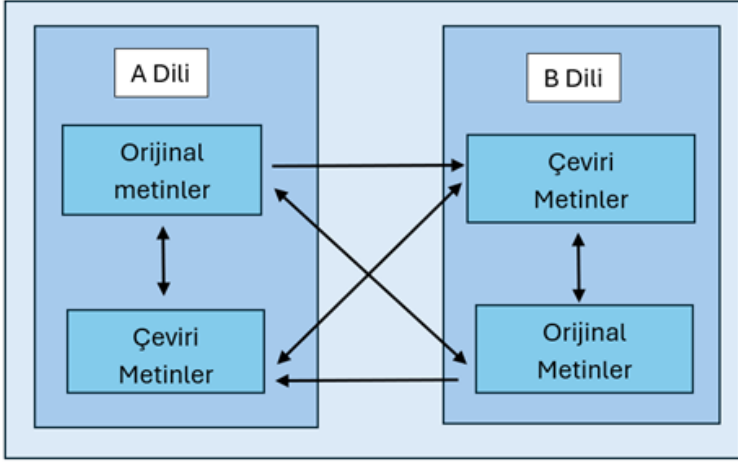
(c) A dilinde oriinal metinler ve A dilinde (farklı) çeviri metinler. (çeviri derlemi)

Farklı adlandırmalar Baker (1995), Borin (2002), Johanson (2002), ve Laviosa (2003) tarafından kullanılmıştır. Karşılaştırma derlemleri ve çeviri derlemleri (a) ve (b) esas itibarıyla benzer yapıdaki tek dilli derlemlerdir ve amaçları ise benzerlikleri ve farkları bulmak için farklı dilleri ve bir dilin farklı lehçelerini karşılaştırmaktır (örn. İngiliz ve Amerikan İngilizcesi). Bu tür derlem verilerini işlemek için standard tek dilli derlemleri analiz etmek için kullanılan yöntemler ve araçlar kullanılır. Diğer taraftan, paralel derlemlerin tenmel amacı bunlardan farklı olarak oriinal metinler ve onların çevirileri arasındaki farkları araştırmaktır.

### 1.3.2. Paralel Derlemler

Paralel bir derlem, orijinal metinler ve diğer dillerdeki çevirileri olmak üzere aynı metinlerin iki veya daha fazla versiyonunu içeren bir derlem olarak tanımlanabilir. İdeal olarak, bir paralel derleme eklenecek çeviri metinlerin profesyonel çevirmenler tarafından gerçek hayatta kullanılmak üzere çevrilmiş metinler olması gerekir, örneğin yayınlanmış kurgu eserlerinin çevirileri, uluslararası internet sitelerinin yerel versiyonları, çok dilli turist broşürleri, ikili ve çok taraflı anlaşmalar vb. Paralel derlemler öğrenciler tarafından yapılan çevirilerden de derlenebilir ve öğrenci derlemi olarak adlandırılırlar. Bu tür derlemlerden çevirmen eğitiminde faydalanılabilir.

Paralel derlemler genellikle,  $A \rightarrow B$  ve  $B \rightarrow A$  olmak üzere her iki yönde çevirilerden oluşur. Bu tür derlemler çeşitli diller arası geçişleri kolaylaştırmak için özel olarak oluşturulmuştur. Büyüklük açısından (her dilde benzer sayıda kelime içerirler) ve içerik olarak (her dil için benzer metin türleri dahil edilmiştir) dengeli bir dağılım gösterirler. Bu şekilde iki yönlü bir derlem kullanılarak orijinal ve çeviri metinleri karşılaştırmanın yanı sıra orijinal metin yazarları tarafından kullanılan dil ile çevirmenler tarafından kullanılan dil karşılaştırılabilir.



Şekil 3: İki yönlü bir derlemede dil ilişkileri. (Baker, 1995)

Paralel derlemler aynı metnin, kaynak dildeki orijinal metin ve hedef dildeki çeviri metin olmak üzere iki dilde bulunduğu bir yapıda olan derlemlerdir. Paralel derlemlerin etkili kullanılabilmesi için hizalanmış olmaları gereklidir. Hizalanmış derlemler kullanılarak farklı dillerdeki karşılıklı cümleler incelenebilir. Cümle hizalaması otomatik çeviri ve dilbilim araştırmalarına olanak sağlar. Hizalanmış derlemler makine çevirisinin geliştirilmesi, yabancı dil eğitimi ve karşılaştırmalı dilbilim gibi alanlardaki çalışmalarda kullanılabilir. Paralel derlemlerin gelişmesi ile birlikte farklı dillerde derlemlerin erişilebilirliği artmıştır ve ilk dönemlerde sadece İngilizce-Fransızca dil çiftleri üzerine yoğunlaşılırken, günümüzde Çince, Lehçe ve Türkçe gibi diller de eklenmiştir.

Paralel derlemler, çeviri çalışmaları, karşılaştırmalı dilbilim ve çok dilli araştırmalar için gereklidir. Araştırmacıların ve çevirmenlerin orijinal metinleri ve çevirilerini karşılaştırmalarına olanak tanıyarak diller arasındaki dilsel olguların incelenmesine yardımcı olurlar. Önemlerine rağmen, paralel

derlemler tek dilli derlemlere göre daha nadirdir ve derlenmesi daha zordur. Günümüzde, çoğu dilbilimsel araştırma elektronik verilere dayanmaktadır. İster teorik dilbilimsel araştırma alanında, ister dilbilgisi ve sözlüklerin derlenmesinde olsun, derlem farklı dillerin yapısını, morfolojisini ve sözdizimini incelemek için standart bir araç haline gelmiştir. Her türden elektronik metin derlemleri – tam metin koleksiyonları, metin örnekleri, kaydedilmiş konuşma transkriptleri vb. oldukça yaygındır ve hatta derlem verilerini kullanmayan araştırmalara şüphe ile yaklaşılır hale gelmiştir.

Birçok dil için 'ulusal derlem' adı verilen derlemler derlenmektedir. Bu eğilim İngiliz Ulusal Derlemi (Burnard, 2000) ile başlamış ve daha sonra Amerikan, Polonya, Çek ve Rus ulusal derlemleriyle devam etmiştir (Ide ve Suderman, 2004; Čermák, 2007; Plungian ve Biryukov, 2008). İnternet tabanlı megaderlemler ise COCA, Sketch Engine ve Aranea gibi derlemlerle temsil edilmektedir (Davies, 2008; Kilgarriff v.d., 2004; Benko, 2014). Uygun metin derlemine erişimi olmayan veya kendi derlemine derlemek istemeyenler için internetin kendisi bir derlem olarak kullanılabilir. Bu nedenle, derlem mevcudiyeti sorunu hala çözülmekten uzak olsa da, tek dilli derlem dilbilimi hızla ilerlemektedir.

Ancak paralel derlemler, daha teknik uygulamalarda, özellikle makine çevirisi alanında, yani otomatik çeviri programlarının geliştirilmesi ve test edilmesinde de çok önemlidir. Paralel derleme ihtiyaç duyulan bir diğer önemli alan, karşılaştırmalı dilbilim olarak bilinen daha teorik alandır. Bu, karşılaştırmalı dilbilgisi ve sözlükleri derlemek amacıyla diller arasındaki morfolojik, sözdizimsel ve sözcüksel benzerlikleri/farklılıkları araştırır. Aynı zamanda, dillerin ortak noktaları olan farklı dil tümellerinin incelenmesiyle de ilgilidir. Ek olarak, paralel derlem kullanılarak yapılan karşılaştırmalı

arařtırmaların sonuçları, dil öğretiminde kullanılan yöntemler ve ders materyalleri üzerinde bir etkiye sahip olacaktır. Gerçekten de paralel derlemeler hem öğretmenler hem de dil öğrenenlerin kendileri tarafından sınıfta bile kullanılabilir.

Öyleyse, paralel derlemin gelişimi neden tek dilli derlemin gerisinde kalmıştır? Bunun nedeni, bir dilde çok sayıda metin elde etmenin, birkaç farklı dilde karşılık gelen versiyonları olan metinleri bulmaktan çok daha kolay olmasıdır. Ayrıca metin hizalama, yani farklı dillerdeki karşılık gelen cümleleri birbirine bağlama sorunu da vardır. Bu nedenle, paralel derlemeleri derlemek zaman alıcı bir girişimdir ve bu gelişimlerinin neden tek dilli derlemesine ayak uyduramadığını açıklar (Salkie, 2002). Fakat paralel derlemeler aynı zamanda daha teknik uygulamalar için de gereklidir. Özellikle makine çevirisi alanında otomatik çeviri programlarının geliştirilmesi ve test edilmesi için kilit rol oynarlar. Paralel derlemelerin bir diğer kullanım alanı da daha teorik bir alan olan karşıtsal dilbilimdir. Karşıtsal dilbilim diller arasındaki biçimbilimsel, sözdizimsel ve sözcüksel benzerlikleri ve farklılıkları inceleyerek karşılaştırmalı dilbilgileri ve sözlükler derlemeyi amaçlamaktadır. Ayrıca, farklı dillerin ortak özelliklerini konu edinen dilsel evrensellerin incelenmesiyle de ilgilenmektedir.

Ek olarak, paralel derlemeler kullanılarak yürütülen karşılaştırmalı arařtırmaların sonuçları, dil öğretiminde kullanılan yöntemleri ve ders materyallerini de etkilemektedir. Örnek olarak, paralel derlemeler hem öğretmenler hem de dil öğrenenler tarafından sınıf ortamında da kullanılabilir. Yabancı dilde yazarken veya çeviri yaparken çok dilli veriye ihtiyaç duyulmaktadır. Terminolojiyi kontrol etmek, uygun deyimsel ifadeler bulmak, bilinen bir alıntının standart veya farklı mevcut çevirilerini tespit etmek ya da bir alıntının orijinalde

nasıl kullanıldığını öğrenmek gerekebilir. Ancak mevcut paralel ve karşılaştırılabilir derlemlerin büyük bir kısmı, boyutlarının yetersizliği veya tam metinler yerine örneklemelerden oluşturulmaları nedeniyle bu amaçlar için kullanılamamaktadır. Teorik olarak, bu eylemlerin birçoğu geleneksel internet aramaları (Google veya diğer ticari arama motorları aracılığıyla) ya da Wikipedia gibi çok dilli kaynaklara başvurarak gerçekleştirilebilir; ancak bu tür çok dilli internet aramaları, yalnızca tek bir dilde yapılan aramalara kıyasla kullanıcının daha yaratıcı olmasını gerektirmektedir.

Benzer şekilde, akademik araştırmalarda, iki veya daha fazla dilin yapılarının incelenmesinde veya iki dilli sözlüklerin hazırlanmasında kullanılan paralel derlemlerin, araştırmacının güvenilir sonuçlara ulaşabilmesi için yeterli büyüklükte olması gerekir. Bununla birlikte, incelenen dillerin yeterli şekilde temsil edilmesini sağlamak amacıyla bu derlemler çeşitli metin türlerini de içermelidir. Birden fazla dilde bu tür metinleri bulmak, tek bir dille çalışmaya kıyasla çok daha zordur. Tüm bu hususlar, paralel derlemlerin tek dilli derlemlere kıyasla neden çok daha az yaygın olduğunu ve paralel derlemlerin sunduğu avantajların neden hâlâ tam olarak takdir edilmediğini açıklamaktadır.

### **1.2.3. Karşılaştırılabilir Derlemler**

Karşılaştırılabilir derlemler, birbirinin çevirisi olmayan ancak temel açılardan mümkün olduğunca benzer olacak şekilde seçilen iki veya daha fazla dilde metin koleksiyonlarıdır. Altenberg ve Granger'ın tanımına göre, karşılaştırılabilir derlemler "her dilde metin türü, konu ve iletişim işlevi açısından mümkün olduğunca eşleşmiş orijinal metinlerden" oluşur (Altenberg ve Granger 2002:7-8). Çeviri derlemleri gibi, benzer iç yapıya sahip tek dilli alt derlemlerden oluşurlar, ancak

amaçları kaynak-hedef ilişkilerini incelemek değildir. Bunun yerine, benzer derlemler, İngiliz İngilizcesi ve Amerikan İngilizcesi gibi dilleri veya dil çeşitlerini karşılaştırmak için kullanılır; böylece sistematik benzerlikler ve kullanım farklılıkları tespit edilir. Sadece orijinal metinleri içerdikleri için, karşılaştırılabilir derlemlere uygulanan analitik araçlar ve yöntemler, standart tek dilli derlemlerde kullanılanlarla büyük ölçüde aynıdır; paralel derlemler ise orijinaler ile çevirileri arasındaki ilişkileri araştırmak için özel olarak tasarlanmıştır (Mikhailov ve Cooper, 2016).

#### 1.2.4. Özel Amaçlı Derlemler

Derlem tabanlı çalışmalar yapmak amacıyla 1960'lı yıllardan itibaren özel konular ile ilgili derlemler derlenmeye başlanmıştır ve bu çalışmalar günümüze kadar devam etmiş ve etmektedir. Özel amaçlarla derlenen derlemler sözcükbilim araştırmaları yapmak, sözlük oluşturmak, konuşma dili üzerine araştırma yapmak, dil edinimi üzerine araştırmalar yapmak, artzamanlı (diachronic) olarak dil değişimini araştırmak gibi farklı konulara odaklanmaktadır. (Kennedy, 1998)

##### *a. Sözcükbilim Derlemleri*

Sözcükbilim üzerine araştırma yapmak amacıyla derlenen ilk derlemler arasında, Algeo'nun (1998) çalışmalarında yararlandığı ve 18. yüzyıldan günümüze uzanan metinleri kapsayan, toplamda yaklaşık beş milyon kelimelik veriye ulaşan Brown derlem ailesi sayılabilir. İlk elektronik derlem örneklerinden olan bu derlemlerin temel amacı, özellikle İngiliz İngilizcesinin sözcüksel özelliklerini incelemektir. Bununla birlikte, bu derlemler artzamanlı olarak İngiliz ve Amerikan İngilizcesi arasında karşılaştırmalı kültürel ve dilbilimsel araştırmalar yapılmasına da olanak sağlamıştır.

Sözcükbilim amaçlı ilk derlem örneklerinden bir diğeri ise *American Heritage Intermediate* (AHI) Derlemidir. (Carroll, Davies, Richman, 1971) AHI derlemi 7-15 yaş arası okul çocuklarına yönelik hazırlanmış okuma kitaplarından oluşan 5,09 milyon kelimelik bir derlem ve her biri 500 kelimelik 10,043 metin örneği içermektedir. AHI derleminin amacı *American Heritage School Dictionary* adlı sözlük için bir referans kaynağı oluşturmaktır ve sözcükbilim araştırmaları için derlenmiş ilk elektronik veri bankalarından biriydi. Bu derlem kendisinden sonra 1980'lerden itibaren yayınlanan *Longman Dictionary of Contemporary English* ve *Collins Cobuild English Language Dictionary* gibi birçok ticari derlem tabanlı sözlük projelerinin öncüsü olması açısından önem taşımaktadır.



Şekil 4: American Heritage Dictionary (<https://ahdictionary.com>)

### b. Sözlük Derlemler

Elektronik derlemler sözlük oluşturmak için kullanılmıştır fakat bunun yanında bazı sözlükler de özel derlemler olarak işlev görmüşlerdir. Özellikle çeviri gibi amaçlar için doğal

dil işleme sistemlerinin geliştirilmesinde sözcüklerin farklı anlamları hakkında detaylı bilgiye ihtiyaç duyulması elektronik sözlüklerin geliştirilmesini gerekli kılmıştır. Derlem olarak kullanılan sözlükler içinde 1980'lerde en fazla tanınan Longman Dictionary of Contemporary English'dir (LDOCE) (Summers, 1987). 1990'larda ise belli başlı standart sözlükler elektronik olarak yayınlanmış ve sözcükbilim çalışmalarına kelime kullanım sıklıklarının da eklenmesiyle olasıksal bir boyut kazandırmıştır.



Şekil 5 : Longman Dictionary of Contemporary English  
(LDOCE, <https://www.ldoceonline.com/>)

Oxford English Dictionary (OED) sözlüğünün CD-ROM olarak yayınlanması bu bağlamda çığır açıcı olmuştur (Simpson ve Weiner, 1989). Sözlük yarım milyon sözcük ve 2.4 milyon açıklayıcı cümle içermektedir. OED aynı zamanda CD-ROM üzerinde kendi detaylı arama yazılımını sunduğu ve sözcükler etiketlenmiş olduğu için yazılım farklı kategorilerdeki her bir kelimeyi rahatlıkla bulabiliyordu. Örneğin, tam olarak sözlük veya derlem olarak adlandırılmasa da Oxford Psycholinguistic Database yaklaşık olarak 99.000 İngilizce

kelime ile ilgili bilgileri bir araya toplayarak derlem dilbilimi arařtırmaları için zengin bir sözcüksel veritabanı saęlıyordu (Coltheart v.d., 1987). Her kelime için 27 farklı özellikle ilgili bilgiler verilmekteydi; bu bilgilerden bazıları kelimenin harf sayısı, sesbirimleri (phonemes) ve heceleri, kelime sınıfı hakkında bilgi, kelimenin diyalektal, soyut veya somut oluşu ile ilgili bilgi, Brown ve Lund derlemlerine göre kelimenin kullanım sıklığı bilgileri şeklinde sıralanabilir (Quinlan, 1992; Kennedy, 1998).

### *c. Sözlü İngilizce Derlemleri*

Sözlü İngilizce derlemleri İngilizcenin sözcüksel ve dilbilgisel incelenmesinin yanı sıra söylem analizi ve bürünbilim (prosody) arařtırmalarının yapılmasına olanak saęlamıştır. Günümüzde birçok dilde sözlü derlemler bulunmaktadır fakat sözlü derlemlerin ilk örnekleri İngilizce dilinden derlenmiştir. İlk sözlü derlemler ve özellikleri ařağıdaki tabloda özetlenmektedir.

Tablo 2. İlk Sözlü Derlem Örnekleri (Kennedy, 1998).

Derlem	Kelime sayısı	İçerik	Kullanım alanları
<b>Lancaster/IBM Spoken English Corpus (SEC)</b> (Knowles et al., 1992)	52,600 kelime	Sözlü standart İngiliz İngilizcesi, 1984-1987 arası, 11 kategori, radyo haberleri, üniversite dersleri, dini yayınlar, kurgu yayınlar, şiir, diyalog ve propaganda metinleri.	Sözcüksel çalışmalar, söylem analizi, bürünsel (prosodic) çalışmalar
<b>London-Lund Corpus</b> (Svatrvik, 1975, 1990)	100 konuşma metni toplam 500.000 kelime	Bürünsel (prosodic) özellikleri açısından etiketlenmiş aynı ve farklı statüden kişiler arasındaki tartışma ve konuşmalar, telefon konuşmaları, hazırlıklı ve hazırlıksız topluluk önünde yapılan konuşmalar, spontane radyo konuşmaları	Sözcüksel, dilbilgisel araştırmalar, söylem çözümlemesi, bürünsel (prosodic) çalışmalar.
<b>Corpus of Spoken American English</b> (Chafe et al., 1991)	80 saatlik konuşma çözümlemesi	900 konuşmacıya ait sözlü dil örnekleri, diyalog, monolog, spontane ve planlı konuşma örnekleri, yüz yüze konuşmalar, telefon konuşmaları, Tartışmalar, hizmet sunumu, satış konuşmaları, işyeri konuşmaları, iş ve politika toplantıları, danışma, ders anlatımı, vaaz, dua, söyleşi programı, haber programları, spor haberleri	Lehçe çalışmaları, konuşma dili ve yazı dili arasında karşılaştırmalı çalışmalar
<b>Map Task Corpus</b> (Anderson et al., 1991)	147,000 kelime	64 Glasgow Üniversitesi öğrencisi arasında 128 iki kişilik karşılıklı konuşma örnekleri. Kontrollü bir deneysel çalışma.	Sözlü dilde değişkenliğin araştırılması.
<b>Wellington Corpus of Spoken New Zealand English</b> (Holmes, 1995)	1 milyon kelime	1988-1993 yılları arasında toplanmış resmi, yarı resmi ve gündelik monolog ve diyaloglardan oluşan konuşma örnekleri.	

## BÖLÜM II - Derlem Dilbilimi ve Çeviribilim

*“Translation is not a matter of words only: it is a matter of making intelligible a whole culture.”*

*Anthony Burgess*

Derlem dilbilim arařtırmalarının bařlangıcı olarak verilebileceđimiz 1950'lerin bařında ampirizm, pozitivizm ve davranıřçılık akımlarının etkisi ile dilbilimciler dil arařtırmalarının otantik metinlere odaklanması üzerinde durmuřlar ve ilk derlemler dil edinimi, dil öğretilimi, sözdizimi anlambilim ve fonoloji alanlarında kullanılmaya bařlanmıřtır. 1950'lerin ortalarında dilbilimde hâkim olan görüşler Noam Chomsky'nin üretici-dönüřümsel dilbilgisi kuramına dayanmaktaydı (Chomsky, 1965).

**Tüm insanlar, farklı ve sonsuz sayıda dilbilgisel cümle üretme yetisini doğuřtan tařır ve mevcut metinlerin analizi, dilbilgisel olmayan yapıları dilbilgisel olanlardan ayırt etme yeterliliđini açıklayamaz (Chomsky, 1965, ss. 3-4).**

Chomsky'ye göre dilbilim, mevcut metinlerde bulduđumuzdan ziyade dil yetisini veya bir konuřmacının yeni dilbilgisel cümleler üretme yeteneđini tanımlamayı amaçlar, çünkü metinler bize yalnızca performans hakkında bilgi verir, yeterlilik hakkında bilgi vermez. 1980'lerden bu yana, dil arařtırmaları ve diđer amaçlar için giderek artan sayıda derlem derlendiđinden, derlem temelli dil arařtırmalarına olan ilgi yeniden canlanmıřtır. Derlem dilbilimcileri Chomsky tarafından savunulan rasyonel yaklařımı, doğrulanamaz ve öznel olduđu konusunda eleřtirmişlerdir. Bunun karřısında, derlem

yaklaşımı, nicel ve nitel yaklaşımların bileşkesini vurgulayarak daha nesnel ve bilimsel bir yaklaşım sunmaktadır. Ayrıca, bilgisayar teknolojisinin giderek yaygınlaşan kullanımı, bilim adamlarının dil araştırmalarında derlemleri daha rahat ve etkili bir şekilde derlemelerine ve kullanmalarına olanak sağlamıştır (Hu ve Kim, 2024).

Derlemlerin sistematik olarak oluşturulması ve araştırılması, betimleyici ve uygulamalı dilbilimde sürekli olarak yeni araştırma alanları açmaktadır. Karşılaştırmalı dilbilim, sözlükbilim, eğitimsel dilbilim, bilişimsel dilbilim, terminoloji, ikinci dil edinimi, adli dilbilim, edebiyat çalışmaları ve eleştirel dilbilim gibi alanlarda, derlem dilbiliminin etkisi sayesinde metodolojik ve teorik yaklaşımlarında önemli değişimler yaşamıştır. Özellikle çeviri çalışmaları'nda, derlem temelli araştırma, çeviriyi kavramsallaştırmak, incelemek ve öğretmek için yeni bir yol ortaya çıkarmıştır. Laviosa (1998), derlem ve çeviribilim arasındaki dayanışmayı çeviribilimin teorisi, betimlemesi ve uygulamasıyla ilgili çeşitli konuları ele alan tutarlı, bileşik ve zengin bir paradigma olarak tanımlamıştır. Kuralcı (Prescriptive) çalışmalar, araştırma hipotezlerinin detaylandırılmasına, analiz yöntemlerine ve çeviri davranışın altında yatan varsayılan dilsel, kültürel ve ideolojik ilkelere odaklanır. Betimsel çalışmalar ise, farklı diller hakkında önemli miktarda karşılaştırılabilir ampirik veri toplar. Diğer taraftan, uygulamalı alan, çevirmen yetiştirmek için yenilikçi ve etkili yollar geliştirir. Bu yeni paradigma, Derlem Tabanlı Çeviri Çalışmaları (DTÇ), çevirinin ürün ve sürecinin ampirik olarak incelenmesi, teorik yapıların detaylandırılması ve çevirmenlerin eğitimi için orijinal ve/veya çeviri metin derlemlerini kullanan disiplinin dalı olarak tanımlanabilir. DTÇ, titiz ve esnek bir metodoloji kullanır, teorik ilkeler sıkı bir şekilde ampirik gözlemlere dayanır, çeviri ve çeviri

arařtırmalarında hem tümevarımsal hem de tündengelimsel yaklařımlar kullanır ve teorik, ampirik ve uygulamalı arařtırmacılar arasında diyalog ve iř birlięini teřvik eder.

1990'lı yılların ortalarından bu yana, derlem yaklařımı eviribilimde yaygın olarak kullanılan derlem yaklařımı ařaęıdaki ilkeleri vurgulamaktaydı:

**evrilmiř metnin nicel ve nitel analizinin birlikte kullanılması.**

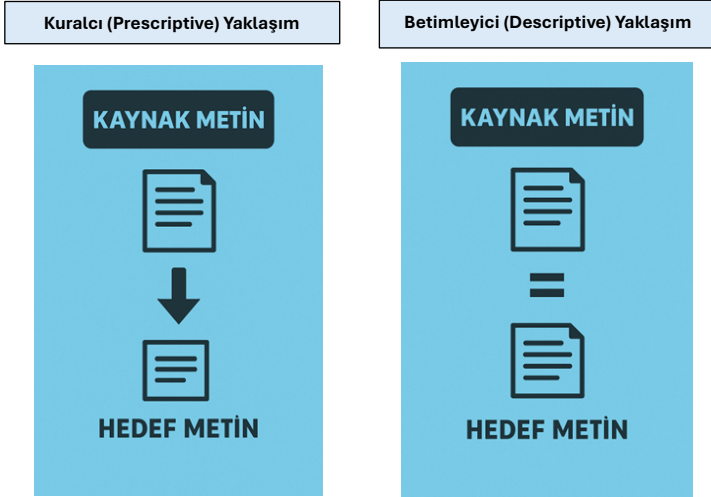
evrilmiř metinlerin zelliklerinin, szdizimsel, szcksel, anlamsal ve metinsel dzeylerde aıklanması.

**eviri evrensellerinin, evirmenin tarzının, eviri normlarının tretilebileceęi, znel yargılar ve olası nyargılar zerine yapılan sonulardan kaınılması.**

Derlem yaklařımı, zgn metinlerin analizi iřıęında belirli bir hipotezin doęrulanmasına vurgu yapan tabandan yukarı iřlelemeyi (bottom-up) temel alan bir yaklařım olarak tanımlanabilir. Uzun bir sre boyunca, eviri dili, geleneksel olmayan ve doęal olmayan bir dil tr olarak kabul edildi ve bu nedenle derlem dilbiliminin nesnesi olarak grlmedi. Ek olarak, 1990'lara kadar eviri alıřmalarında kaynak dil odaklı, kuralcı yaklařım baskın olmuřtur. Bu paradigmaya gre hedef metin, kaynak metnin trevidir ve bu nedenle, evirinin doęası veya srecinden baęımsız olarak, kaynak ve hedef metinler arasındaki eřdeęerlięe odaklanılır ve hedef metnin kaynak metne eřdeęer olması beklenir. Bu yaklařımda, evirinin zelliklerine, hedef metnin hedef kltrdeki rolne ve eviriyi yneten eřitli faktrlere daha az dikkat edilir. Bu, tercme edilmiř metinlerin sistematik ve kapsamlı bir Őekilde arařtırılmasına yol amıřtır (Baker, 1993; Toury, 1995; Holmes, 1988; Laviosa, 1998).

## 2.1. Betimleyici Çeviri Çalışmaları ve Derlem Dilbilimi

Betimleyici Çeviri Çalışmaları (Descriptive Translation Studies - DTS) bağlamında ise kaynak metin ile hedef metin arasındaki ilişkiye farklı bir akışı geliştirilmiştir. Baker (1993), Gideon Toury'nin başlattığı betimleyici çeviri çalışmaları geleniğini geliştirmiş ve “iyi” ya da “kötü” çeviriden ziyade, çeviri ürünlerinde gözlemlenen eğilimleri ve stratejileri sistematik olarak tanımlama amacı gütmüştür. Betimleyici yaklaşımda, kaynak ve hedef metinler arasındaki ilişki eşdeğerlik (equivalence) kavramı üzerinden değil, metinlerarası düzenlilikler (regularities) üzerinden incelenir. Baker'a (1993) göre hedef metin, kaynak metnin birebir yeniden üretimi değil; sistematik düzenliliklere sahip bir yeniden yazım sürecidir. Kaynak metin ve hedef metin arasındaki ilişki, metinler arası uyum (alignment), dilsel seçimler ve bağlam üzerinden değerlendirilir. Bu yaklaşım, çeviri çalışmalarında kuralcı (prescriptive) modellerden uzaklaşarak ampirik bir bakış açısını savunur.



Şekil 6. Kuralcı (Prescriptive) ve Betimleyici (Descriptive) çeviri yaklaşımlarının karşılaştırılması (Toury, 1995).

Baker, 1990'larda derlem dilbilimi yöntemlerini çeviri çalışmalarına entegre eder. Büyük ölçekli paralel ve karşılaştırmalı derlemeler aracılığıyla KM (Kaynak Metin) ve HM (Hedef Metin) arasındaki ilişkiler istatistiksel ve ampirik olarak araştırılır. Baker'a (1993) göre, derlem temelli çeviri çalışmaları, çeviri metinlerinde ortaya çıkan ve çeviri sürecinde yer alan belirli dil çiftlerinin etkisinden bağımsız dilsel özellikler olarak tanımlanabilecek "dolaylı bir iletişimsel olay olarak çevrilen metnin doğasını" ve çevirinin "tümellerini" ortaya çıkarması açısından önemlidir. Bu bağlamda, çeviri metinler gerçek iletişimsel olayları kaydeder ve bu nedenle herhangi bir dildeki diğer iletişimsel olaylardan ne aşağı seviyededir ne de üstündür. Bu yaklaşım, çeviri metinlerde ortaya çıkan tipik özellikleri, yani çeviri evrensellerini (translation universals) tanımlar:

**Sadelik (Simplification):** HM'de dilin daha açık ve anlaşılır hâle gelmesi.

**Açıklık (Explicitness):** Bilgilerin HM'de daha belirgin hâle gelmesi.

**Normalleştirme (Normalization):** HM'de hedef dilin normlarına daha fazla uyum sağlanması.

**Müdahale (Interference):** KM'nin yapısının HM'de iz bırakması (Baker, 1993).

Derlem dilbilimi ve çeviribilim alanlarının paylaştığı ortak yaklaşımlar vardır. Her iki alandaki araştırmalar da idealleştirilmiş veya sezgisel dil verilerine değil doğal olarak meydana gelen dil kullanımı örneklerine odaklanır. Ayrıca, her iki yaklaşımda da verilerde gözlemlenen dilsel kalıplar, değişmez kurallar olarak değil, olasılıksal davranış normları olarak

kabul edilir. Bu dil kalıpları, kültürü yansıttıkları ve yeniden ürettikleri ölçüde sosyokültürel değişkenlerle ayrılmaz bir şekilde ilişkili kabul edilir. Her ikisi de ampirik kanıtlarla desteklenen tanımlayıcı araştırmanın önemini ve bağlamsallaştırmanın gerekliliğini vurgular.

Derlem temelli çalışmalarda kullanılan farklı derlem türleri farklı amaçlara hizmet eder. Çeviribilimde kullanılan üç tür derlem vardır: paralel derlem, çok dilli derlem ve karşılaştırılabilir derlem. Paralel derlem çevirmen eğitiminde ve makine çevirisinde kullanılabilir ve kuralcı çeviri araştırmasından betimleyici çeviri araştırmasına geçişi mümkün kılar. Çok dilli derlem ise farklı çeviri metinlerin stillerini tanımlamak için kullanılabilir ve böylece farklı çeviri metinler arasında karşılaştırma yapmak için yararlı parametreler sağlayabilir. Bu bağlamda derlem çeviribilimde bir çevirmenin çeviri tümellerinin özellikleri veya "tipik olarak orijinal ifadelerden ziyade çeviri metinde ortaya çıkan ve belirli dilsel sistemlerden gelen müdahalenin sonucu olmayan" özelliklerin araştırılmasında kullanılabilir (Baker 1993, s. 243).

Baker (1999) Derlem Dilbilimi ve Çeviri çalışmalarında kullanılan terminoloji arasındaki farklılıklara odaklanarak bu farkları açıklamıştır. Örneğin, derlem dilbilimi terminolojisinde karşılaştırılabilir derlemler genellikle çok dilli (faklı dillerde karşılaştırılabilecek metinler) iken, çeviri çalışmalarında genellikle tekdillidirler (aynı dilde çeviri ve orijinal metinler). Çeviri çalışmaları bağlamında Malmkjaer'in (1998) tanımına göre paralel derlem genellikle, karşılık gelen çevirileri ile hizalanmış bir dizi kaynak metinden oluşan derlemler olarak düşünülür, fakat bu tür derlemler karşıt dilbilimciler tarafından çeviri derlemi olarak adlandırılır. Baker (1995) ortak bir terminolojinin gerekliliğini vurgular ve bir tipoloji öne sürer.

Bu tipolojide çokdilli ve tekdilli derlemler arasında ayrıma gidilir.

**Çokdilli derlemlerin iki türü vardır:**

(a)Çeviri Derlemi (veya paralel derlem): Bu derlemler kaynak metinleri ve çevirilerini içerir ve tek yönlü (X dilinden Z diline) veya iki/çok yönlü olabilirler.

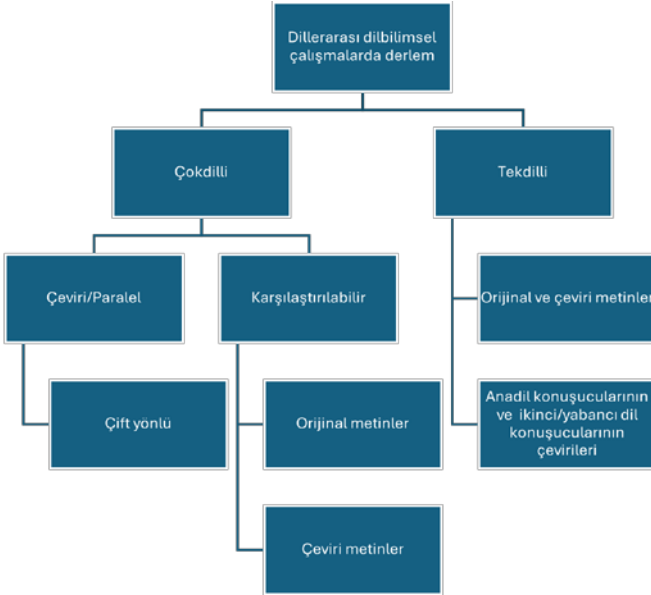
**(b) Karşılaştırılabilir Derlemler: Bu derlemler aynı türden çevrilmemiş veya çevrilmiş metinler içerir.**

•Baker (1995)

Tek dilli Derlemler: Karşıt dilbilim araştırmaları ile ilgili tüm tek dilli derlemlerin tümü karşılaştırılabilir derlemlerdir. Bunlar ise aşağıdakileri içerebilir.

- Bir veya aynı dilde orijinal ve çeviri metinler,
- Bir ve aynı dilde anadili ve öğrenci metinleri

Bu tipolojide paralel derlem terimi kaynak metinler ve onların çevirileri olarak tanımlanır. Karşılaştırılabilir derlemlere atıfta bulunmak için veya herhangi birçok dilli derlem türüne atıfta bulunmak için genel bir terim olarak kullanılmaz (Teubert 1996, s.245).



Şekil 7: Dillerarası Dilbilimsel Çalışmalarda Derlem (Granger vd. 2008)

Bu tipolojideki farklı derlem türleri kullanılarak çeşitli karşılaştırmalar yapılabilir.

	Karşılaştırma Türü	Derlem Türü	Alan
1	ODx ↔ ODy	Orijinal metin derlemlerinin çokdilli karşılaştırılması	KD
2	KDx ↔ ÇDy	Çokdilli çeviri / paralel derlem	KD ve ÇB
3	KDx ↔ ÇDy	Orijinal ve çeviri metinlerden oluşan tekdilli karşılaştırılabilir derlem	ÇB ve KD
4	ÇDx ↔ ÇDx	Çeviri metinlerden oluşan çokdilli karşılaştırılabilir derlem	ÇB

OD: orijinal dil / KD: kaynak dil / ÇD: çeviri dil / ÇB: Çeviribilim / KD: Karşılaştırmalı Dilbilim

Şekil 8: Çeviribilim ve derlem dilbilimi bağlamında derlem türleri ve karşılaştırma ilişkileri (Baker, 1995).

Laviosa'ya (1998) göre Derlem Tabanlı Çeviri çalışmaları ile Betimleyici Çeviri Çalışmaları arasındaki önemli bir ilişki bulunmaktadır. 80'lerde çeviribilimde betimleyici yaklaşımı geliştiren Gideon Toury olmuştur. Toury'nin (1995) Holmes'un (1972) temel haritasından uyarlanan programında, Betimleyici Çeviri Çalışmaları (BÇÇ) alanın kalbinde yer alır: kendine özgü bir iç organizasyonu vardır, çeviri teorisiyle etkileşime girer ve Toury'nin artık uygulamalı Çeviri Çalışmaları olarak adlandırmadığı, ancak uygulamalı uzantılar olarak adlandırdığı şeyle tek yönlü bir ilişkisi vardır.

Toury'nin BÇÇ için öne sürdüğü metodolojik prosedürlerin amacı işlev, ürün ve çeviri süreci arasındaki ilişkileri ortaya çıkarmaktır ve metot çalışma nesnesinin tanımlanmasıyla başlar. Burada çalışmanın nesnesi, 'hangi gerekçeyle olursa olsun, hedef kültür içinde sunulan veya bu şekilde kabul edilen tüm ifadeler' olarak tanımlanan 'varsayılan çevirilerden' oluşur (Toury 1995, s.32). Bu tanımın gerekçesi, BBC'nin dayandığı iki temel ilkedен kaynaklanmaktadır. Birincisi, "çeviriler hedef kültürlerin gerçekleridir: bazen özel bir statüye sahip, hatta bazen kendi içinde tanımlanabilir (alt) sistemlerini oluşturabilirler, ancak her durumda hedef kültürün gerçekleridirler" (Toury 1995, s. 29). İkincisi, çevirilerin sadece diğer metinlerin temsilleri değil, kendi başlarına metinler olduğunu onaylar. Bu hedef odaklı ve ampirik bakış açısı içinde, tek tek metinleri veya bir metin derlemine seçmek için ilk kriterler dışsal, geçici ve hedef dil sistemine sıkı sıkıya bağlıdır.

Toury'nin BÇÇ yaklaşımı üç temel aşamadan oluşur. Varsayılan çevirilerin analizinin ilk aşaması kaynak metne atıfta bulunulmadan, her bir metnin hem genel olarak hedef dil metni olarak hem de hedef kültürde bir çeviri olarak kabul edilebilirliğinin değerlendirilmesini içerir. Analizin ikinci

aşaması, Toury tarafından önerildiği gibi, kaynak metnin tanımlanmasıyla başlar ve bir hedef metni kaynağına bağlayan ilişkileri belirlemek için hedef metnin bölümlerinin kaynak metindeki karşılıkları ile eşleştirilmesi ile devam eder. Metin bölümlerinin paralel analizi, çeviride kaynak metne göre meydana gelen kaymaları incelemek için de kullanılabilir. Analizin üçüncü aşamasında, kaynak ve hedef metinler arasında kurulan ilişkiler, seçilen metin çifti için eşdeğerliği yöneten norm hakkında genellemelerin temelini oluşturur.

## 2.2. Veriye Dayalı Çeviribilim

Baker'a (1995) göre çeviribilimde derlem tabanlı diğer bir deyişle veri tabanlı bir yaklaşım kullanılarak çeviri üzerine farklı araştırma konuları üzerine araştırma yapılabilir. Bunlar aşağıda sıralanmıştır:

**Çeviri sürecinin nihai ürününün zaman içinde gelişimi.**

- Çeviri biriminin büyüklüğü ve niteliği ve uygulamadaki eşdeğerlik türleri.

**Metin türlerinin rolü, çeviri türleri ve iletişimde yer alan türlerin ve söylemlerin çeşitliliği.**

- Baker (1995)

Derlem araştırmalarında bahsedilen ilk derlem tabanlı araştırma çabası, Oléron ve Nanpon'un 1965'deki çalışmasıdır. Yeni milenyumdan önce gerçekleştirilen en az yedi proje daha incelemelere dahil edilmiştir, ancak bu tür erken derlemelerin makine tarafından okunabilir olması pek olası değildir ve muhtemelen derlem dilbilimsel araçları bu çalışmaların

bir parçası değildi (Setton, 2011). Fakat buna rağmen derlem olarak sınıflandırılmışlardır çünkü Setton'un çalışması gerçek hayattaki çeviri görevlerinden (anekdotlar, iç gözlem veya deneyler değil) elde edilen ampirik veriler anlamına gelen "otantik derlem" (Setton 2011, s. 38) üzerine odaklanır. Çalışmada listelenen diğer ilk derlem projeleri ise artan sayıda tamamen makine tarafından okunabilir transkriptlere dayanmaktadır (ve bazı durumlarda bunlar etiketlenmiş ve dizine eklenmiş şekildedir) ve oluşumların otomatik olarak çıkarılmasından yararlanır, bu nedenle derlemin genel tanımı olan " belirli bir dizi kritere göre elektronik biçimde bir araya getirilmiş geniş bir otantik metin koleksiyonu" ile tamamen uyumludur ve " (Bowker ve Pearson 2002, s. 9). Bununla birlikte, bireysel araştırmacılar tarafından manuel analiz için geçici olarak oluşturulan derlemlerin yorumlanması bugün hala mevcuttur ve Derlem Tabanlı Çeviri alanını tamamlar. Sözlü Çeviri Çalışmalarında derlem kavramının başlangıçta gerçek verilere (yani gerçek hayattan sözlü çeviri grevlerine) dayanan ampirik araştırmalarla bağlantılı olduğu ve elektronik derlem oluşturmadaki zorlukların bu kavramı manuel olarak analiz edilen veri setlerine uygulanabilir kıldığı açıktır. Nitekim doksanlı yıllarda dilbilimciler tarafından önerilen çeşitli derlem tanımlarına bakıldığında, özgünlük, temsil edilebilirlik ve örnekleme kriterlerini ana noktalar olarak ön plana çıkarsa da "otomatik veya yarı otomatik işleme veya analize uygun, bilgisayarlı bir otantik metinler koleksiyonuna" atıfta bulunan bir tanım yapılmıştır (Tognini-Bonelli, 2001).

Tüm öncü çalışmalar, konferans ortamlarında gerçekleştirilen profesyonel simultane çeviriye odaklanmıştır. Bu eğilim yıllar boyunca sabit kalmıştır ve özellikle iki özel veri kaynağıyla, yani TV yayıncılığı ve Avrupa Parlamentosu (AP) ile desteklenmiştir. Sonuncusu, özellikle, EPIC (Monti vd., 2005)

projesi gibi büyük derlem tabanlı çeviri projeleri geliştirme fırsatı vermiştir. Aynı anda 23 farklı dile çevrilen kaynak ve hedef konuşmaların bol miktarda bulunması ve verilerin araştırma amacıyla kullanımına izin verilmesi sayesinde çok sayıda araştırma yapılabilmektedir (Bendazzoli, 2010). AP verilerinin çekiciliğine rağmen, festivaller, tıbbi konferanslar, futbol basın toplantıları ve diğerleri gibi daha fazla konferans ortamı araştırılmıştır. Öte yandan, Asya araştırma merkezlerinde yapılan çalışmalarda ardıl çeviri çalışmalarının, simultane çeviri çalışmalarına göre daha fazla sayıda olduğu görülmektedir. Bunun nedeni, bir kez daha veri erişilebilirliği ile bağlantılıdır. Aslında, Ardıl Çeviri ile ilgili tüm araştırma projeleri aynı veri kaynağına, yani Çinli siyasi temsilcilerin televizyonda yayınlanan basın toplantılarına dayanmaktadır (Wang 2015). Buna ek olarak aynı zamanda topluluk ortamlarında kısa ardıl çeviri veya diyalog çevirisi ile ilgili projeler de mevcuttur ve gizlilik ve anonimlik sorunları nedeniyle bu tür verileri toplamanın önünde daha büyük engeller olmasına rağmen, işaret dili çeviri derlemi geliştirmek için umut verici çabalar da gösterilmektedir (Metzger ve Roy, 2011).

Çeviri Derlemleri, İngiliz Ulusal Derlemi'nin (British National Corpus) sözel kısmı gibi genel referans derlemleriyle karşılaştırıldığında, boyut açısından hala oldukça sınırlıdır (Dembry ve Love 2015). AP verilerine dayanan projeler bile henüz genel referans derlemlerinin boyutuna ulaşmamıştır, ancak AP bağlamında simultane çeviri hakkında genelleme yapmak için yeterince büyük kaynaklara ulaşma potansiyeli olduğu için bu sadece bir zaman ve işgücü meselesi olabilir. Bu kaynaktan veri toplamanın avantajlarına rağmen, ortam türlerini çeşitlendirmek ve sonuçların temsil gücünü geliştirmek için diğer uluslararası kuruluşların Derlem Tabanlı Çeviri araştırmasının nesnesi haline gelmesi umulmaktadır.

Güven verici bir şekilde, Avrupa Komisyonu (Scardulla, 2016) ve Birleşmiş Milletler (Dayter, 2016) gibi diğer uluslararası kuruluşların da dikkate değer çalışmaları vardır. Sözlü Çeviri ile ilgili literatüre göre (Matsubara vd. 2002; Ono ve ark. 2008; Tohyama et al. 2005), Nagoya Üniversitesi'nin Simultane Çeviri Veritabanı (SIDB) veya Setton'un çalışmasında bildirildiği gibi CIAIR, bugüne kadar bir milyon kelime içeren tek çeviri derlemidir.

# BÖLÜM III - ÇEVİRİBİLİM ARAŞTIRMALARI VE UYGULAMALARINDA DERLEM KULLANIMI

*“Regularities emerge where attention is systematic.”*

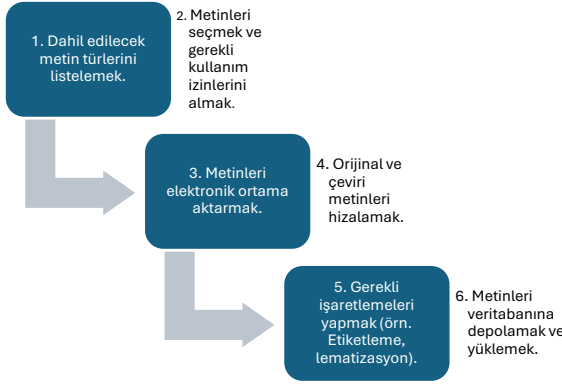
Douglas Biber

## 3.1. Çeviribilim Odaklı Derlem Tasarımı

Çeviribilim arařtırmalarını destekleyebilecek birçok derlem mevcuttur, fakat bazen bu derlemler arařtırmacıların üzerinde durdukları özel arařtırma sorularını cevaplamada yetersiz kalabilirler. Bu durumda arařtırmacıların kendi derlemlerini oluřturmaları gerekebilir. Bu tür derlemler DIY derlemler (do-it-yourself-corpora) olarak adlandırılır ve genellikle arařtırmacılar tarafından oluřturulan küçük ölçekli derlemlerdir (Mahadi v.d. 2010; Zanettin 2012). Bu derlemlere farklı arařtırmacılar tarafından farklı isimler de verilmiřtir. Örneğın Varantola (2003) ‘disposable corpora’ ve Pastor ve Seghiri (2009) ‘virtual corpora’ kavramlarını kullanmıřlardır.

## 3.2. Planlama Adımları

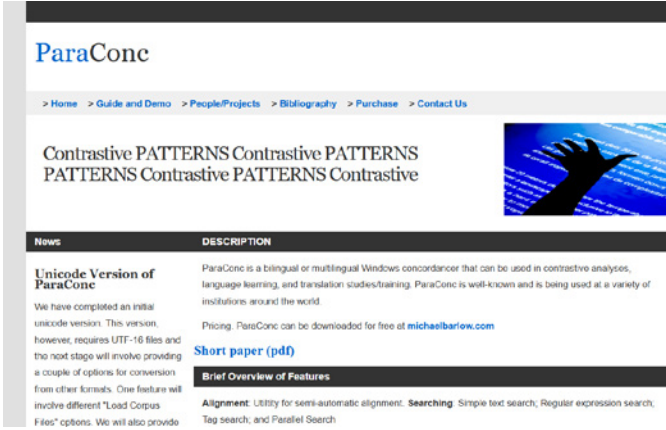
Çeviribilim arařtırmalarında kullanılmak üzere tasarlanan bir paralel derlem oluřturmak için izlenecek adımlar ařağıdaki şekilde sıralanabilir.



Derlem planlamasının ilk aşamalarında dikkat edilmesi gereken bazı noktalar vardır. İlk olarak, derlemenin her aşamasında örneğin taramada, hizalamada ve işaretlemelerde kullanılacak yazılımların seçimi dikkatlice yapılmalıdır. Basılı metinler ile çalışılacak ise bir OCR (optik karakter tanıma) yazılımı kullanılarak metinler taranmalı ve elektronik ortama depolanmalıdır. Çevrimiçi metinler kullanılacak ise kod biçimlendirme ve temizlik yapmak gerekecektir ve bu durumda temel programlama becerileri faydalıdır.

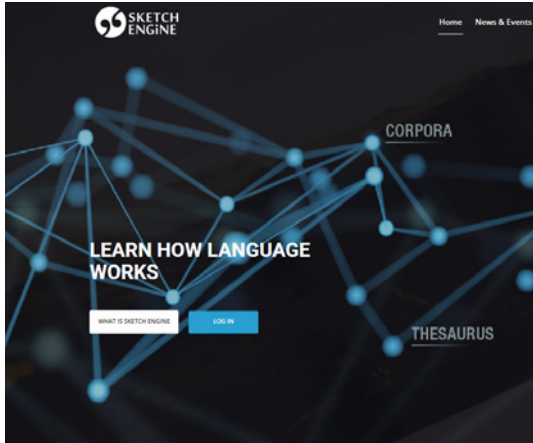
#### *a. Yazılım Seçimi*

Kullanılabilecek faydalı yazılımlar arasında hizalama yazılımları, lematizasyon ve POS etiketleme (sözcük türü etiketleme) yazılımları ve depolama ve arama sistemleri sıralanabilir. Bir derlem ideal olarak bir arama arayüzü, bir bağlam dizini aracı gibi istatistiksel araçlar ve bir güncelleme ve bakım yazılımı içermelidir. Çoğu derlem yazılımı tek dilli derlemler için tasarlandığından hizalanmış paralel derlemler için uygun değildir.



Şekil 9: ParaConc (<https://paraconc.com>)

Windows ve Linux işletim sistemleri için mevcut programlar bulunmaktadır. Bunlar arasında Windows işletim sisteminde ParaConc (Barlow, 2003) ve SketchEngine (Kilgarriff v.d., 2004) gibi yazılımlar paralel metinlerle çalışmak için kullanılabilir. Linux sistemi için ise IMS Corpus Workbench (Evert ve Hardie, 2011) örnek verilebilir.



Şekil 10: SketchEngine (<https://www.sketchengine.eu>)

Alternatif olarak programlama bilgisine sahip olan arařtırmacılar Perl, PHP veya Python gibi programlama dillerini kullanarak kendi derlem yazılımlarını yazabilirler. Örneđin WordSmith araçları amatör bir proje olarak başlamıřtır (McEnery ve Hardie, 2012). Ayrıca R yazılımı veri iřleme ve istatistiksel analiz için kullanılabilir; arama yapmak ve istatistiksel testler yürütmek, özet tablolar, grafikler ve histogramlar oluşturmak için kullanılabilir (Gries, 2009).

### *b. Derlem Büyüklüğüne Karar Verilmesi*

Bir derlemin ilk planlama aşamalarından biri olan derlem büyüklüğüne karar vermek mevcut zaman, personel ve teknik sınırlamalar ile yakından bağlantılıdır. Derlem büyüklüğüne karar verilirken öncelikle derlemin yapısı ve içeriđi konusunda kararlar almak gerekir. Örneđin derleme dahil edilecek metinler örnekleme řeklinde birçok metinden alınan alıntılar mı yoksa tam metinler mi olacaktır. Örnekleme derlemi oluşturmak daha uzun zaman alabilir ama temsil etme gücü daha yüksektir. Buna örnek olarak Brown Derlemi (referans) gösterilebilir (Francis ve Kućera, 1964). Bir milyon kelimeden oluşan Brown Derleminin planlanması ve derlenmesi 3 yıl sürmüřtür. Paralel derlem oluşturulurken manuel örnek metin seçimi yapılması gerekli olabilir, çünkü otomatik rastgele seçimler hizalanmıř derlemlerde iyi sonuç vermeyebilir. Görece küçük örnek derlemler bile dikkatlice tasarlanırsa büyük olanlardan daha güvenilir ve temsil gücü yüksek olabilir.



kaynak hem de hedef derlem metinlerinin ait oldukları dili ve kültürleri yeterince temsil ettiklerinden emin olmak gerekir. Tek dilli derlemelerdeki bazı metinlerin çevirileri mevcut olmayabilir. Metin seçimleri yapılırken buna da dikkat edilmelidir (Bowker & Pearson, 2002; Zanettin, 2012). Temsil edilebilirliği sağlamak için öncelikle kaynak metin alt-derlemi planlanmalı, sonrasında çevirilerin varlığı ve kalitesi kontrol edilmelidir. Derlemede dengenin sağlanması için de seçimler bilinçli olarak yapılmalıdır.

#### *e. Metin Girişlerinin Yapılması*

Bir derleme metinlerin eklenmesi basit ve rutin bir iş değildir. Metinler halihazırda dijital olsa bile derlem araştırmasına uygun olup olmadıklarından emin olmak için dikkatlice kontrol edilmeli, temizlenmeli ve biçimlendirilmelidir. Derleme metin eklenirken metinlerin özelliklerine göre farklı yöntemler izlenmesi gerekebilir. Metinlerin manuel olarak yazılması, taranması veya OCR (Optik Karakter Tanıma) kullanılarak dijitalleştirilmesi gerekebilir.

Bu süreçte özel olarak dikkat edilmesi gereken unsurlar vardır:

- Yazım hataları ve OCR hataları düzeltilmelidir. (örneğin “rn” OCR tarafından “m” olarak yanlış okunabilir.
- Farklı diller ve alfabeler (örneğin, Kiril, Yunanca, İskandinav harfleri ä, ø, å) kodlama sorunlarına yol açabilir.
- En iyi uyumluluk için düz metin formatı (UTF-8) kullanılmalıdır (Mikhailov, M., & Cooper, R. (2016).

#### **f. Metinlerin Hizalanması**

Hizalama (alignment), bilgisayarın diller arasında hangi parçaların karşılık geldiğini tanımasını sağlar. Bu, dilbilimsel araştırmalarda ve çeviri çalışmalarında ve çeviri hafızalarının oluşturulması için gereklidir. Hizalamanın manuel olarak

yapılması uzun metinler için zaman alıcı olacaktır. Bu sebeple, genellikle otomatik hizalama yazılımları tercih edilir.

Hizalama yazılımları çevirilerin orijinal metinler ile aynı uzunlukta olduklarını varsayar ve ortalama segment uzunluklarının oranını hesaplar (örneğin İngilizce-Fransızca cümle uzunlukları). Ayrıca formatlama, sayılar ve özel isimleri kontrol ederek eşleşen segmentleri bulurlar. Yine de hatalar olabileceği için her zaman manuel kontrol gereklidir. Bazı metinler (örneğin şiirler, şarkı sözleri) cümle cümle hizalanmaz. Bunun yerine bir şiirdeki kıtaları hizalamak veya Word’de tablolar kullanmak gerekebilir. Bu durumda kaynak ve çeviri kıtalarını satırlar halinde yerleştirmek ve satırlar eşleşene kadar elle hizalamak gerekebilir.

**Box 2.1 A sample of an aligned Russian-Finnish text in TMX format**

```
<tu tuid="000000003">
<tuv xml:lang="ru">
<seg>Но и мамаша его, судя по всему, тоже была из ежиной породы –
она его и не трогала, даже руки ему не протягивала на крутой
тропинке, когда они поднимались от пляжа к дому.</seg>
</tuv>
<tuv xml:lang="fi">
<seg>Siilien heimoon kuului tosin kaikesta päätelten myös lapsen äiti, joka ei
koskenutkaan poikaan eikä edes ojentanut tälle kättä heidän noustessaan jyrkkää
polkua uimarannalta talolle.</seg>
</tuv>
</tu>
```

Şekil 12: TMX formatında kaydedilmiş Rusça-Fince hizalanmış bir örnek (Mikhailov ve Cooper, 2016).

Hizalanmış metinleri kaydetmek için farklı yöntemler kullanılabilir. Örneğin ayrı dosyalar halinde kaydetmek, burada her segmentin aynı kimlik numarasını paylaşması gerekir. Her iki metnin birlikte bir dosyada saklandığı bitext formatı kullanılabilir (örneğin TMX) ve XML tabanlı çeviri belgesi olarak kullanılır veya segment başlangıcı ve sonu gibi

ayrı hizalama bilgilerinin ayrı saklandığı ayrı hizalama dosyaları kullanılabilir (Bowker ve Pearson, 2002).

### **g. Metin Anotasyonu**

Metin anotasyonunda metinlere üstveri (metadata) veya dilbilimsel detaylar gibi ek bilgiler eklenir. Anotasyon ham metinleri, dilbilimsel analiz, çeviri araştırması ve hesaplamalı işleme için faydalı olan zengin ve yapılandırılmış verilere dönüştürür. Üstveri derlemleri aranabilir, analiz edilebilir ve yeniden kullanılabilir hale getirir. Bunun yanında TEI veya XCES gibi ortak standart formatların kullanılması araştırmacıların derlem araştırmalarında ortak çalışma yapabilmelerini sağlar.

Anotasyonda derlem metinlerine iki çeşit ek bilgi eklenir.

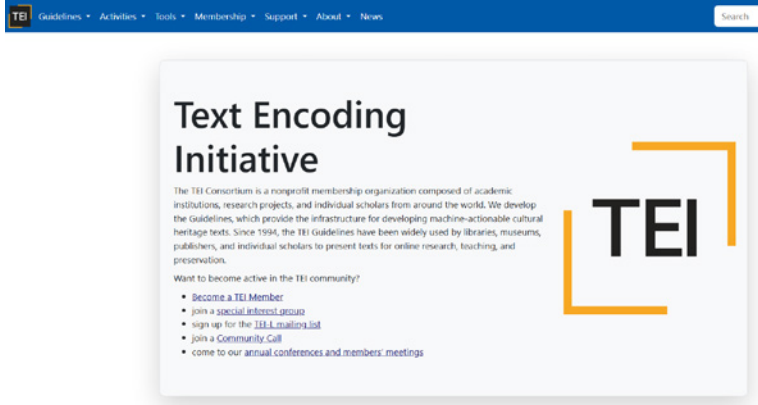
a) Dış Bilgiler (Üstveri): Bu bilgiler yazar, başlık, çevirmen, editör, yayınevi, yıl, dil gibi bibliyografik verileri içerir. Bunların yanında diğer faydalı üstveriler de eklenebilir; örneğin, konu, tür, orijinal ve çeviri, yazarın cinsiyeti/yaşı/anadili gibi. Zengin üstveriler alt-derlemleri tanımlamaya yardımcı olur (örneğin, kadın yazarlar, teknik metinler, orijinal İngilizce).

b) İç Bilgiler (Dilbilimsel Anotasyon)

Dilbilimsel anotasyon metnin içine farklı işaretlemeler eklenmesini içerir. Bu işaretlemeler derlemin türü ve amacına göre metnin yapısı, hizalama, pragmatik bilgiler, dilbilgisel bilgiler ve fonetik bilgilerden oluşabilir. Yapısal bilgilerle ilgili anotasyonlarla metin bölümleri, paragraflar ve cümleler işaretlenir. Hizalama ile ilgili anotasyonda paralel derlemlerdeki kaynak metin hedef metin ilişkileri ile ilgili bilgiler eklenir. Pragmatik anotasyonda konuşmacı, alıcı, söylem işlevi gibi bilgiler eklenir. Sözcük-dilbilgisel anotasyonda ise sözlükbirim (lemma), sözcük sınıfı (POS), sözdizim ve anlam ile

ilgili bilgiler eklenir. Konuşma derlemlerinde bunlara ek olarak fonetik bilgi de eklenir (Mikhailov ve Cooper, 2016).

Derlem dilbiliminde anotasyon uygulamalarında tutarlılık, birlikte çalışabilirlik ve bilimsel verilerin yeniden kullanılabilirliğinin sağlanması için büyük ölçüde standart çerçevelerin kullanılmasına dayanır. En etkili standartlardan biri, çalışma alanları arasında metin materyallerinin kodlanması ve işaretlenmesi için uluslararası geçerliliği olan yönergeler sunan TEI'dir (Metin Kodlama Girişimi) (Bowker ve Pearson, 2002). TEI, metinlerin yapısal, dilsel ve yorumlayıcı özelliklerini temsil etmek için esnek ama titiz bir çerçeve sunar. Bu ilkeler üzerine inşa edilen 'Corpus Encoding Standard for XML' (XCES), TEI yönergelerini özellikle derlem yapısı için uyarlar ve sistematik dilsel açıklama ve hesaplamalı işlemeyi kolaylaştırmak için XML tabanlı bir yapı kullanır.



Şekil 13: TEI-Text Encoding Initiative (<https://tei-c.org>)

Derlem anotasyonunda, hem manuel hem de otomatik işaretleme yaklaşımları belirgin avantajlar ve sınırlamalar sunar. Manuel işaretleme, potansiyel olarak kesin olsa da, zaman alıcı yapısı ve sonraki aşamalarda anotasyonların doğrulanması, düzeltilmesi veya sürdürülebilmesinin zorluğu

nedeniyle genellikle tavsiye edilmez. Ayırıştırma ve etiketleme araçları aracılığıyla uygulanan otomatik işaretleme, daha yüksek verimlilik sunar ancak anotasyon hatalarına yol açabilir ve ardından manuel inceleme gerekebilir.

Özellikle sözcük sınıfı etiketleme (POS: Part of Speech Tagging) için çeşitli etiketleme araçları geliştirilmiştir. İngilizce için yaygın olarak kullanılan sistemler arasında British National Corpus'ta (Burnard, 2000) kullanılan Biber Tagger (Biber v.d., 1998) ve genel amaçlı etiketleme için yaygın olarak kullanılan Stanford Tagger (Toutanova v.d., 2003) bulunur. Bu etiketleyicilerden Biber Tagger (Bkz. Ek B), metinlerdeki dilbilgisi ve söylem özelliklerini otomatik olarak tanımlamak ve etiketlemek için tasarlanmış bir dilsel açıklama aracıdır. Douglas Biber tarafından geliştirilen çok boyutlu analiz çerçevesine dayanır; bu çerçeve, izole unsurlar yerine birlikte ortaya çıkan dilsel özelliklerin kalıplarını yakalamayı amaçlamaktadır. Bu çerçeve, birden fazla fonksiyonel boyut boyunca metinlerin ve metin türlerinin sistematik olarak karşılaştırılmasını mümkün kılar. Etiketleyici, konuşma bölümleri, zaman ve aspekt biçimleri, zamirler, pasifler, bağlantılar, duruş işaretçileri ve diğer söylemle ilgili yapılar dahil olmak üzere geniş bir dilsel özellik yelpazesini notlar. Bu açıklamalar, dilsel kalıpların nicelendirilmesini ve metin türleri veya orijinal ile çevirilmiş metinler arasında karşılaştırmalı analizler yapılmasını mümkün kılar. Bu nedenle, Biber Tagger derlem dilbilimi, söylem analizi ve derlem tabanlı çeviri çalışmalarında yaygın olarak kullanılmıştır. Biber Tagger'ın ana güçlü yanlarından biri, çıktısının çok boyutlu istatistiksel analizi desteklemek için özel olarak tasarlanmış olmasıdır. Bu, araştırmacıların metin türleri arasındaki sistematik farklılıkları belirlemesine ve dilsel özelliklerin farklı iletişim bağlamlarında nasıl bir arada toplandığını keşfetmesine olanak tanır. Bu araç,

büyük ölçekli derlemler için uygundur ve kontrast, kayıt tabanlı ve diyakronik derlemaraştırmaları için sağlam bir metodolojik temel sağlar. Ayrıca, birçok Avrupa dili için birkaç ticari olmayan etiketleyici mevcuttur. Bu sistemlerin çoğu manuel olarak etiketlenmiş örneklerden öğrenerek ve olasılıksal modeller kullanarak çalışan Brill etiketleyici (Brill, 1995) gibi istatistiksel etiketleyicilerdir.

Derlemlerin temel iki türü vardır; bunlar etiketlenmiş (annotated) ve etiketlenmemiş (un annotated) derlemelerdir. Etiketlenmemiş derlemlerde herhangi bir dilbilimsel açıklama eklenmemiş düz metinler yer alır. Etiketlenmiş derlemlerde ise metinlere dilbilimsel özelliklerine göre açıklamalar eklenir. Etiketlenmiş derlemler örtük bilgilerin açık olarak ulaşılabilir olmasını sağlayarak derlem araştırmalarında kullanım kolaylığı sağlar. Örneğin, etiketlenmemiş bir derlemde yer alan '*loves*' (sevmek yüklemine geniş zaman hali) kelimesi herhangi bir etiket eklenmeden yer alırken, etiketlenmiş bir derlemde kelimenin yanına onun sözcük türünü gösteren bir etiket eklenir ve kelime derlemde '*loves\_WZ*' olarak yer alır. Burada '*\_WZ*' etiketi kelimenin türünü gösterir. Etiketler etiketleme yapılacak olan dilin dilbilim kuralları gözetilerek belirlenmiş tutarlı bir etiket listesine göre eklenir. Etiketleme günümüzde özel olarak tasarlanmış yazılımlar tarafından yüksek güvenilirliğe sahip olarak bilgisayar ortamında yapılabilmektedir.



English-Corpora.org

corpora PDF guides videos related resources users my account [upgrade](#) [help](#)

Overview: [brief](#) | [detailed](#) [+](#) Now available: AI/LLM insights integrated into corpus results

These are the most widely used online corpora, and they serve many different purposes for teachers and researchers at universities throughout the world. In addition, the corpus data (e.g. full-text, word frequency) has been employed by a wide range of companies in many different fields, especially technology and language learning.

The links below are for the free online interface. You can also purchase and download [1](#) the corpora for use on your own computer.

Corpus	Download	# words	Dialect	Time period	Genre(s)
<a href="#">News on the Web (NOW)</a>	<a href="#">1</a>	23.7 billion+	20 countries	2010 yesterday	Web: News
<a href="#">iWeb: The Intelligent Web-based Corpus</a>	<a href="#">1</a>	14 billion	6 countries	2017	Web
<a href="#">Global Web-Based English (GloWbE)</a>	<a href="#">1</a>	1.9 billion	20 countries	2012-2013	Web (incl blogs)
<a href="#">Wikipedia Corpus</a>	<a href="#">1</a>	1.9 billion	(Various)	2014	Wikipedia
<a href="#">Coronavirus Corpus</a>	<a href="#">1</a>	1.5 billion	20 countries	2020-2023	Web: News
<a href="#">Corpus of Contemporary American English (COCA)</a>	<a href="#">1</a>	1.0 billion	American	1990-2019	Balanced
<a href="#">Corpus of Historical American English (COHA)</a>	<a href="#">1</a>	475 million	American	1820-2019	Balanced
<a href="#">The TV Corpus</a>	<a href="#">1</a>	325 million	6 countries	1950-2018	TV shows
<a href="#">The Movie Corpus</a>	<a href="#">1</a>	200 million	6 countries	1930-2018	Movies
<a href="#">Corpus of American Soap Operas</a>	<a href="#">1</a>	100 million	American	2001-2012	TV shows

Şekil 14: British National Corpus (<https://www.english-corpora.org>)

Etiketleme dilbilimsel analizlerin daha hızlı ve doğru şekilde yapılmasına olanak sağlar ve metinleri yapılandırılmış bir dilbilimsel veri bankasına dönüştürür. Etiketlenmiş derlemeler dilbilgisi, sözdizim ve anlambilim alanlarında farklı dilbilimsel araştırmalar için faydalıdır. Etiketleme araştırmalara güvenilir bir altyapı oluşturabilmesi için sistemli olarak belirli ilkeler çerçevesinde yapılmalıdır. Bu bağlamda Leech (1997) etiketleme için 7 kural ileri sürmektedir. Bunlar aşağıdaki şekilde listelenebilir:

## Etiketleme Kuralları

1. **Çıkarılabilir olma:** Etiketler istendiğinde metinlerden çıkarılabilmeli ve ham metinler elde edilebilmelidir.
2. **Ayrıştırılabilir olma:** Etiketler istendiğinde bağımsız olarak depolanabilmek için metinlerden ayrıştırılabilmelidir.
3. **Yönergeye dayalı olma:** Etiketleme, tutarlılığı sağlamak için önceden belirlenmiş yönergeleri takip etmelidir.
4. **Şeffaflık:** Etiketlemenin nasıl ve kim tarafından yapıldığı konusundaki detaylar açık ve net olmalıdır.
5. **Yorumlanabilir olma:** Etiketlemeler katı değil yoruma açık olmalıdır.
6. **Teori tarafsızlığı:** Etiketleme genel olarak kabul gören dilbilimsel ilkelere bağlı olmalıdır.
7. **Katı standartların olmaması:** Standartlar uygulamadaki fikir birliği ile belirlenir.

Etiketlemede bazı zorluklarla karşılaşabilmektedir; örneğin, kullanıcı ihtiyaçları ile etiketleme zorluğunun dengelenmesi, büyük ölçekli etiketlemelerde hatasızlık ile pratikliğin dengelenmesi gibi. Örneğin bazı kelimelerin anlamsal ayrımları otomatik etiketleme için fazla ayrıntılı olabilmektedir.

Derlem etiketlemede bazı standart formatlar kullanılmaktadır:

*COCOA referansları:*

COCOA referansları, metinlerle ilgili bilgilerin açılı parantez içinde sunulduğu bir metaveri kodlama formatıdır (<>). Bu sistem, derlem analiz programlarının bağlamsal ve bibliyografik bilgileri ana metinden ayrı olarak tanımlaması ve

işlemesi için geliştirilmiştir. COCOA formatında, yazar, başlık, tarih veya tür gibi meta veriler açıkça işaretlenmiş etiketlerle doğrudan metin dosyasına gömülür. Örneğin, yazar bilgisi <CHARLES DICKENS> şeklinde üst veri olarak kodlanabilir. Bu referanslar, metnin dilsel içeriğinin bir parçası değildir; bunun yerine, araştırmacıların bir derlem içindeki metinleri filtrelemesini, sınıflandırmasını ve karşılaştırmasını sağlayan yapısal işaretleyiciler olarak işlev görürler. COCOA referansları, özellikle erken derlem dilbiliminde etkili olmuştur; çünkü dilbilimsel verileri betimleyici bilgilerle ilişkilendirmek için standartlaştırılmış ve makine tarafından okunabilir bir yol sunarak, sonraki derlemlerde kullanılan daha gelişmiş notlama ve işaretleme sistemlerinin yolunu açmıştır.

*Metin Kodlama Girişimi (TEI, Text Encoding Initiative):*

Diğer bir etiketleme formatı olan Metin Kodlama Girişimi (TEI: Text Encoding Initiative) ise Standart Genelleştirilmiş Biçimlendirme Dili'ne (SGML: Standard generalized Markup Language) dayanmaktadır (Burnard & Bauman, 2014). TEI formatında metin yapısal olarak metaverileri tanımlayan Başlık ve ana metni içeren Metin olarak ikiye ayrılır. TEI formatında ayrıca Metin Türü Tanımlayıcıları (DTD: Document Type Descriptions) da yer alır. DTD bir metin içindeki bileşenleri tanımlar (örn. Tümceleri paragraflar) ve TEI kapsamında farklı metin tiplerinin standartlaşmasını sağlar (Sperberg-McQueen & Burnard, 2002). DTD uygulamalarına örnek olarak British National Derlemi ve Lampeter Derlemi gösterilebilir (Burnard, 2000; Smith & Baker, 2006). TEI-LITE ise TEI formatının kullanım kolaylığı sağlamak amacıyla geliştirilmiş daha basitleştirilmiş bir alt türüdür. TEI' ye uyumlu olmakla birlikte daha az etiketleme unsuru içerir ve yeni başlayanlar ve küçük ölçekli projeler için çözüm sunar (Burnard & Sperberg-McQueen, 2002).

### *EAGLES ve Derlem Kodlama Standardı (CES)*

Corpus Encoding Standard(CES): EAGLES dil etiketleme standartları için Avrupa Birliği'nin danışma kuruludur ve birçok farklı dil için sözcük türü ve sözdizimsel etiketleme şemaları tanımlar (EAGLES, 1996). CES ise Avrupa Birliği dillerinin nasıl etiketleneceğini tanımlar ve TEI ve SGML ile uyumluluğu öngörür (Ide & Véronis, 1995).

### *Ayrık Etiketleme (Standoff Annotation) ve XML*

Ayrık etiketleme metni etiketlerden ayıran bir sistemdir ve etiketleri doğrudan metne gömmek yerine XML kullanarak kelime pozisyonlarına referans verir (Ide, Nancy ve Romary, Laurent, 2004). Bu da metnin okunmasında ve işlenmesinde kolaylık sağlar. İngilizce olmayan metinlerin etiketlenmesinde birtakım zorluklar bulunmaktadır. Aksan işaretleri ve Romen alfabesi dışındaki alfabe sistemlerini kullanan Yunanca, Rusça, Japonca, Çince ve Arapça gibi dillerin etiketlenmesinde sorunlar yaşanabilmektedir. Buna ek olarak IBM uyumlu kişisel bilgisayarlardaki 8-bitlik karakter setleri de zorluklar yaratabilmektedir. Bu durumlarda alternatif etiketleme yöntemleri kullanmak gerekmektedir. Bu yöntemler arasında Fransızca dili için aksan işaretlerinin kaldırılması örnek olarak verilebilir. Almanca dili için ise standart olmayan örneğin 'Frühling' kelimesindeki 'ü' harfi yerine 'ue' yazılarak kelime 'Fruehling' olarak kullanılabilir veya aynı kelime için tırnak işaretleri kullanılarak kelime "Fr"uhling" olarak kullanılabilir. Yine Almanca dilindeki 'ß' harfi yerine 'ss' kullanılabilir. Fakat farklı kodlama yöntemlerinde veri kayıpları yaşanabilir. Romen alfabesi dışındaki alfabe sistemlerini kullanan diller için ise farklı yöntemler önerilebilir. Bunlar arasında Unicode kullanmak, ISO-646 temelli yapay alfabeler kullanmak, Yunanca harfler yerine onları karşılayacak Romen alfabesi karakterleri kullanmak veya farklı grafik özellikler

elde edebilmek için özel karakter setleri kullanmak sayılabilir (Unicode Consortium, 2023).

Konuşma dilinin etiketlenmesi de beraberinde kendine özgü zorluklar getirmektedir. Bunlar arasında konuşmada belirgin noktalama işaretlerinin olmaması sayılabilir. Bu tümce ve tonlama birimlerinin belirlenmesini zorlaştırır. Bu zorluğa yanıt olarak London-Lund derleminde tonlamaya dayalı deşifre yöntemi kullanılmıştır (Svartvik & Quirk, 1980). Lancaster/IBM Sözlü Derleminde ise yazımsal deşifre yöntemi tercih edilmiştir (Knowles, 1996). Sözlü derlemlerin etiketlenmesinde zorluk yaratan bir diğer unsur da birden fazla konuşmacının aynı anda konuştuğu durumlardır. Bu durumda kodlamada Stave sistemi kullanılarak örtüşen konuşmalar için paralel satırlar kullanılabilir (Edwards, Lampert, Martin , 1993). TEI sisteminde ise örtüşmeler “\*\*” sembolleri kullanılarak gösterilmektedir (TEI Consortium, 2023). Konuşma dilinin etiketlenmesi ile ilgili bir başka zorluk da tümcelerdeki hatalı başlangıçları, duraksamaları ve sözcüksel olmayan sesleri etiketlemektir. Ayrıca konuşmacıların vücut dili, konuşmacılar arasındaki etkileşim gibi unsurların etiketlenmesidir. Bu durumlarda yorum ekleme veya sınırlayıcılar kullanmak gibi çözümler önerilmektedir.

#### *Dilbilimsel Anotasyon*

Sözcük türüne göre etiketleme derlemdeki her bir kelimeye gramatik kategoriler atanması anlamına gelir. Bu tür etiketleme derlem analizinde, aramaların özelleştirilmesinde ve belirsizliklerin giderilmesinde önem taşır. Sözcük türüne göre etiketleme sistemlerine örnek olarak LOB derleminde kullanılan C1 etiket seti, Sözlü İngilizce Derleminde kullanılan C7 etiket seti veya British National derleminde kullanılan C5 etiket seti gösterilebilir.

### *Kök Çözümleme (Lemmatization)*

Kök çözümleme kelimeleri sözlüksel şekillerine dönüştürme olarak tanımlanabilir ve kelimelerin hem düzenli hem de düzensiz hallerini kapsar. Örneğin İngilizce'deki 'go' (gitmek) fiilinin çekimleri olan 'gone, going, went'. Kök çözümleme derlem temelli araştırmalarda araştırmacıların kelime kullanımlarını etkili bir şekilde analiz etmelerine, sözcük bilgisi ve kelime araştırmalarına ve sıklık ve dağılım analizi yapılmasına yardımcı olur. Kök analizinin otomatik olarak yapılabilmesi için yazılımlar geliştirilmiştir. Bu yazılımlardan birisi olan Beale erişilebilir derlemlerde yaygın olarak kullanılmamaktadır (Leech vd., 1995; McEnery ve Hardie, 2012). Kök çözümlemesi yapılmış derlemlerde örnek olarak SUSANNE derlemi gösterilebilir. Bu derlemde kök çözümleri yapılmış kelime formları yer almaktadır, aynı zamanda kelimelere paralel olarak sözcük türü ve sözdizimsel bilgiler verilmektedir.

Kök çözümlemesi sunan diğer derlemler arasında Lancaster Üniversitesi, Universidad Autónoma de Madrid ve C2V Paris işbirliği ile hazırlanmış olan ve çokdilli derlem araştırmalarına olanak sağlayan CRATER Derlemi (İngilizce, Fransızca, İspanyolca) yer almaktadır (Burnard ve McEnery, 1998). Kök çözümlemesi içeren verilerin yapısında sırasıyla metin referansı, sözcük türü, kelimenin kendisi, kelime kökü ve sözdizimsel inceleme yer alır. Kök çözümlemesine eşlik eden sözdizimsel inceleme kelimelerin sözdizimsel ilişkilerine göre düzenlenmesini içerir. Temelde kullanılan inceleme yöntemi, öbek yapısı dilbilimine dayanır ve tam inceleme ve iskelet inceleme olarak iki çeşidi vardır. Sözdizimsel incelemede kullanılan yöntemlerden biri ağaç yapılı treebank representation' gösterimidir. Dilbilimde ağaç yapısı gösterimi,

sözdizimsel veya anlamsal cümle yapısına açıklama ekleyen ayrıştırılmış bir metin derlemi olarak ifade edilebilir.

Bir derlemede anlambilimsel özelliklerin etiketlenmesinde kelimelerin üslenebilecekleri farklı anlamlar ve içerik analizindeki tarihsel artalanlarına odaklanılır. Örneğin belirli temalarla ilişkili kelime grupları (örn., yaşam). Anlambilimsel etiketlemede kelimeler üzerine evrensel bir uzlaşma olması sebebiyle zorluklar yaşanabilmektedir.

#### **h. Derlemin Depolanması**

Derlemler kullanıcı ihtiyaçları ve Teknik altyapıya bağlı olarak çeşitli erişim formatları ile depolanabilir. Yaygın seçenekler arasında BNCweb gibi web tabanlı hizmetler veya masaüstü arayüzlü yerel kurulumlar bulunur (Hoffmann et al., 2008). Ayrıca, bazı platformlar – özellikle BYU-BNC- birden fazla derlemi tek bir arayüzde entegre ederek daha verimli erişim, karşılaştırma ve büyük ölçekli dilsel analizleri kolaylaştırır (Davies, 2004).

Derlemlerin klasörler halinde düzenlenmiş düz metin dosyaları olarak saklanması, bireysel yüksek lisans tezleri gibi küçük ölçekli projeler için yeterli olabilir; ancak bu yaklaşım daha büyük araştırma ortamlarında sorunlu hale gelir. Ortak veya uzun vadeli projelerde, verinin en güncel versiyonunu takip etmek genellikle zordur ve harici depolama cihazları aracılığıyla kontrolsüz veri dolaşımı riski artar. Ayrıca, yerel dosya depolamaya bağımlılık, farklı sistemler arasında yazılım uyumluluk sorunlarına yol açabilir. Bu sebeplerle, güvenli ve daha yapılandırılmış depolama sistemleri kullanılması önerilir.

Derlem depolama ve yönetimi için en etkili çözüm, derlemin merkezi bir sunucuda bulunduğu ve web arayüzü veya SSH (Secure Shell) gibi güvenli bağlantılarla erişildiği sunucu

tabanlı depolamadır. Bu yaklaşım, veri koruması ve sürüm kontrolünü sağlarken dünya çapında erişim sağlar. Ayrıca, sunucu tabanlı sistemler genellikle sadece bir web tarayıcısı ve gerektiğinde SSH istemcisi gerektirdiğinden kullanıcılar için teknik engelleri azaltır; böylece yerel yazılım kurulumuna gerek kalmaz (McEnery ve Hardie, 2012; Kilgarriff vd., 2014).

### **i. Telif Hakkı**

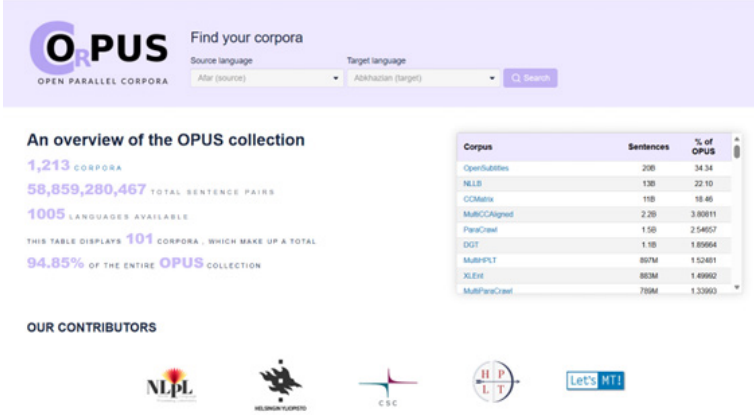
Derlemlerin oluşturulması ve yayılması önemli hukuki ve etik konuları gündeme getirir. Yayınlanan metinler fikri mülkiyet içerdiğinden araştırmacılar derlem oluşturma ve kullanımının telif hakkı düzenlemelerine uygun olmasını sağlamalıdır. Bireysel veya ticari olmayan amaçlarla oluşturulan özel veya küçük ölçekli araştırma derlemlerine genellikle izin verilirken, kamuya açık derlemler genellikle telif hakkı sahiplerinden açık bir yetki gerektirir. Her durumda, derlem tasarımcıları kaynaklarının ticari yayın faaliyetleri ile rekabet etmediğinden ve tam telif hakkı korumalı eserlerin sunulmadığından emin olmalıdır (Mikhailov ve Cooper, 2016).

Özellikle paralel derlemler derlenirken hem orijinal metnin yazarından hem de çevirmenden onay alınması gerekir. Bu tür izinler genellikle ticari olmayan akademik kullanım için verilir ancak resmi olarak imzalanmamış olsa bile genellikle yeterli kabul edilir. Ayrıca derlemlerin kendisi fikri mülkiyet teşkil eder. Bir derlem, derleyicinin açık izni olmadan kopyalanamaz veya yeniden dağıtılamaz; bu da dikkatli dokümantasyon ve mülkiyet haklarına saygı gerekliliğini vurgular.

### 3.3. Çeviribilimde Derlem Analizi için Araçlar ve Teknikler

Derlemlerin çeviribilim odaklı analizinde bazı temel araçlar ve arama yöntemleri kullanılmaktadır. Derlem yazılımları, büyük metin derlemlerine erişmek, arama yapmak ve analiz etmek için tasarlanmış bilgisayar programlarını ifade eder. Çoğu çevrimiçi derlem, sonuçları bağlam dizini çıktısı formunda gösteren bir arama motoru olarak işlev gören ve kullanım sıklığı listeleri ve temel istatistiksel analizler oluşturmak için ek araçlar içeren bir bağlam dizini yazılımı ile donatılmıştır.

Birçok derlem sisteminde, kullanıcıların derlem verilerine doğrudan erişimi yoktur. Veriler serbestçe açılmaz, görüntülenemez veya ham formda indirilemez. Bu kısıtlama, telif haklarını korumak ve derlemin iç yapısını korumak amacıyla tasarlanmıştır. Bunun yerine, kullanıcılar verilerle bağlam dizini gibi arama sonuçları aracılığıyla etkileşime girerler. Önemli bir istisna, telif hakkı olmayan veya kamuya açık materyallere dayalı derlemler için geçerlidir; bunlar düz metin, XML veya TMX gibi formatlarla dağıtılabilir. Bu formatlar özellikle çeviri çalışmaları ve paralel derlem analizi için önemlidir. OPUS derlemi (Tiedemann, 2016), akademik kullanım için indirilebilir paralel derlemler sağlayan bilinen bir örnektir.



Şekil 15: OPUS Derlemi (<https://opus.nlpl.eu>)

Bir derlem araştırma amaçlı olarak indirildiğinde veya derlendiğinde, genellikle entegre bir arama yüzü yoktur ve kullanıcıların analiz için harici yazılım araçlarına güvenmesi gerekir. Yerel olarak depolanan paralel derlemlerde, farklı platformlar özel çözümler sunar. 'ParaConc' (Barlow, 2003) yaygın olarak Windows sistemlerinde kullanılırken, 'IMS Corpus Workbench' (Evert ve Hardie, 2011) Linux tabanlı sistemlerde derlem sorgulama için güçlü bir ortam sağlar. Bu araçlar ham derlem dosyalarında bulunmayan bağlam dizini ve diğer derlem analizi türlerini mümkün kılar.

### 3.1.1. Derlem Üzerinde Arama Yapmak

Bir derlem ile arama yapmak, internette arama yapmak gibi görülebilir; ancak Google gibi genel arama motorları dilbilimsel araştırmalar için tasarlanmamıştır. Ticari arama motorları genellikle tanımlık ve edat gibi işlevsel kelimeleri görmezden gelir, yalnızca yaklaşık kullanım sıklığı bilgisi sunar, tekrarlayan ve kararsız sonuçlar içerir ve birçok dilbilimsel

sorgu türünü desteklemez. En önemlisi arama sonuçlarını sistematik derlem tabanlı analiz için gerekli olan indirilebilir bir bağlam dizini şeklinde sunar.

Derlemlere özel tasarlanmış arama motorları ise genellikle kullanıcıların bir kelime veya ifadeyi – yani arama dizisini – sorgu arayüzüne girmesine izin vererek çalışır; ardından sistem, derlemdeki tüm tam eşleşmeleri elde eder. Bazı araçlar ayrıca özel isimleri dahil etmek veya hariç tutmak ve sadece büyük harfle fark eden biçimleri ayırt etmek için büyük/küçük harf duyarlı aramaları destekler.

Birçok derlem sorgu sistemi, kullanıcıların karmaşık dil sel aramalar yapmasına olanak tanıyan gelişmiş arama fonksiyonlarını destekler. Bunlar arasında ‘wildcard’ olarak adlandırılan joker karakterlerin kullanımı da vardır. Bu joker karakterlerden ‘?’ herhangi bir karakter ve ‘\*’ ise herhangi bir karakter dizisini sorgular. Sistemler ayrıca küme gösterimi kullanılarak alternatiflerin gruplanmasına izin verebilir, böylece aynı anda birden fazla ifade kombinasyonu aranabilir. Ayrıca POS etiketli derlemler, dilbilgisel etiketlere dayalı aramalara izin verir ve kullanıcıların yalnızca yüzeysel kelime biçimlerine dayanmak yerine yapısal olarak tanımlanmış kalıpları, örneğin edat, tanımlık ve seçilmiş sözcük öğeleri kombinasyonlarını elde etmesini sağlar (Mikhailov ve Cooper, 2016).

### *Gelişmiş Derlem Araması Örneği*

#### **Amaç:**

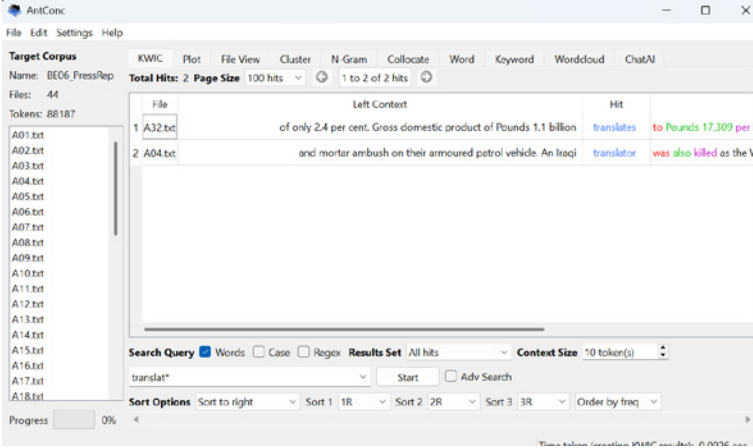
Edat ve isim cümleciklerini içeren yaygın cümle kalıplarını (örneğin sonda, başta tanımlamak)

#### **Adım 1: Joker Karakterler (Wildcard) Kullanmak**

Bir kullanıcı joker karakterler kullanarak bir kelimenin morfolojik varyantlarını arayabilir:

- s?z şeklindeki bir arama söz, saz, süz gibi sonuçlar verecektir.

- transl\* şeklindeki bir arama ise translate, translates, translated, translating gibi sonuçlar verecektir.



Şekil 16: AntConc wildcard arama sonucu

## Adım 2: Alternatifleri gruplandırmak

Alternatif arama komutları, tek bir sorguda birden fazla ifade türünü aramak için kullanılabilir.

{ "bir" | "bu" } {"dönemin"} {"başında" | "sonunda"} Not: {}  
→ sözcük biçimini belirtir  
| → alternatif (veya) operatörü

Yukarıda verilen arama komutu *bir dönemin başında, bir dönemin sonunda, bu dönemin başında, bu dönemin sonunda* gibi sonuçlar verecektir.

### Adım 3: Dilbilgisi etiketleri ile arama

POS etiketli bir derlemde, aramalar dilbilgisel kategoriler kullanılarak formüle edilebilir.

{PREP} {ART} {start | middle} not: {PREP} → edatlar (in, at, on, from vb.), {ART} → tanımlıklar (the, a, an), {} içinde | → alternatif arama

Yukarıdaki sorgu, sözcüklerin biçiminden bağımsız olarak *at the start*, *in the middle*, *from the start*, *to the middle* gibi yapısal olarak tanımlanmış kalıpları verecektir.

Gelişmiş sorgular, araştırmacıların basit kelime aramalarının ötesine geçip derlem verilerindeki tekrarlayan dilbilgisi ve ifadesel kalıpları incelemesini sağlar.

#### *Farklı Arama Sistemleri*

Farklı derlemler farklı sorgu dillerine ve arama sistemlerine dayanır. Örneğin, Corpus Sorgu Dili (CQL) British National Corpus için SARA arayüzünde kullanılırken (Burnard ve McEnery, 2000), Corpus Sorgu İşlemcisi (CQP) Sketch Engine, Aranea ve IMS Corpus Workbench gibi platformlarda kullanılmaktadır. CQP, düzenli ifadeler, sözcük türü etiketleri ve mantıksal operatörler aracılığıyla gelişmiş sorgulamayı destekler (Christ vd., 1999; Evert & Hardie, 2011).

#### **Örnek CQP Sorgusu:**

**Amaç:** Sözcük türü etiketli (POS tagged) bir derlemde sözcük türü etiketlerini, sözcükbirimleri (lemma) ve basit operatörler kullanarak yaygın dilbilgisi kalıplarını elde etmek.

### 1) Herhangi bir sıfat + isim dizisi bul:

Sorgu: [pos="J.\*"] [pos="N.\*"]

Bu sorgu, bir sıfatın hemen ardından gelen bir ismi bulur. Örneğin: *important decision, new policy, economic growth, major problem*. Bu tür aramalar, ad öbeği yapılarının incelenmesinde ve sıfatların isimlerle olan kullanım örüntülerinin belirlenmesinde sıklıkla kullanılır.

### 2) Go kelimesini sözlükbirim olarak bütün çekimleri ile bul (örneğin, *go, went, gone, going*).

Sorgu: [lemma="go"]

**Beklenen sorgu sonucu:** Yüzey biçimi ne olursa olsun *go* sözcükbiriminin bütün çekimlerini bulur.

**İpucu:** Bu sorgu sadece derlem sözcükbirimsel olarak etiketlenmiş ise işe yarar.

### 3) EDAT+TANIMLIK (end/beginning) kalıbını bul:

Sorgu: [pos="PREP.\*"] [pos="ART.\*"]  
[lemma="end|beginning"]

**Beklenen sorgu sonucu:** 'in the end' veya 'at the beginning' gibi dizileri bulur.

**Not:** '|' sözcükbirim dizisinde "veya" anlamına gelir.

### 4) Joker karakter içeren bir sözcük sınıfı (POS) koşulu ekle, herhangi bir sıfat + isim:

Sorgu: [pos="ADJ.\*"] [pos="N.\*"]

**Beklenen sorgu sonucu:** Sıfat-isim kombinasyonlarını bulur (örneğin, *new book, political system*).

5) İkinci kelimeyi isteğe bağlı kılın (örneğin, the end/the very end)

Sorgu: [word="the"] ([pos="ADV.\*"])? [lemma="end"]

**Beklenen sorgu sonucu: 'the end' ve ayrıca 'the very end' gibi bir zarf eklenmiş yapıları bulur.**

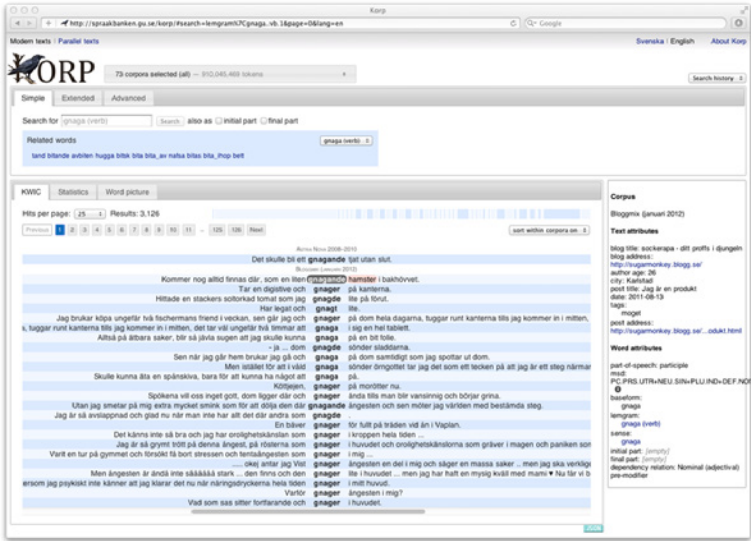
Not: '?' gruplanmış öğeyi isteğe bağlı olarak işaretler.

Not: Tam etiket isimleri (örneğin, N, NN, ART, DET) derlem etiket setine bağlıdır.

Gelişmiş arama araçları, daha verimli ve sistematik veri erişimini mümkün kılarak derlem tabanlı araştırmada kritik bir rol oynar. Karmaşık arama seçenekleri, araştırmacıların zaman kazanmalarını, birden fazla dilsel kalıbı aynı anda tanımlamalarını ve bireysel kelime formlarını ayrı ayrı arama ihtiyacını ortadan kaldırmasını sağlar. Sonuç olarak, gelişmiş sorgulama daha kapsamlı analizleri destekler ve manuel olarak incelenmesi pratik olmayan büyük derlemler üzerinde arama yapılmasını kolaylaştırır.

Çok dilli ve paralel derlem araştırmalarında, arama kaynak metinlerde, ilgili çevirilerde veya bazı sistemlerde her ikisinde aynı anda yapılabilir. Birçok derlem sorgulama aracı, aramaları bir dille sınırlar, bu da karşılaştırmalı ve çeviri odaklı analizleri sınırlayabilir. Önemli bir istisna, Göteborg Üniversitesi'nce geliştirilen ve son derece gelişmiş iki dilli arama kombinasyonlarını destekleyen ve daha esnek diller arası araştırma imkânı sağlayan bir derlem olan KORP'tur (Borin vd., 2012). Korp, İsveç Dil Bankası Språkbanken tarafından geliştirilen bir derlem altyapısıdır. Üç ana bileşenden oluşur: bir derlem işleme hattı (corpus processing pipeline), bir arka plan sistemi (backend system) ve bir kullanıcı arayüzü. Derlem işleme hattı, derlemi içe aktarmaktan, dilsel açıklamalar eklemekten ve etiketlenmiş verileri çeşitli

formatlarda dışı aktarmaktan sorumludur. Bu işleme hattının temel özelliklerinden biri, mevcut yapısal ve kelime düzeyindeki açıklamaları koruyabilmek ve üzerine yeni açıklamalar inşa edebilmesidir. Arka uç bileşeni, kullanıcıların derlemleri aramasına ve ilgili bilgileri almasına olanak tanıyan REST tabanlı web hizmetleri seti sunar. Kullanıcı arayüzü, arka plan sistemi ile iletişim kuran ve kullanıcıların derlemi etkileşimli olarak keşfetmesini sağlayan grafiksel bir arama arayüzüdür. Tasarımı, SketchEngine, Glossa ve DeepDict gibi bilinen derlem araçlarından esinlenmiştir ve State Chart XML (SCXML) kullanılarak kullanıcı etkileşimlerinin yer imlerine eklenmesini destekler. Sistem, ana işlevleri ve teknik yapısına genel bir bakış sunar, ardından gelecekteki gelişim planlarının kısa bir tartışması yapılır.



Şekil 17: KORP Bağlamda Anahtar Kelime Dizini (KWIC) Görünümü (Borin vd., 2012)

Derlem analizinde, sorgu sonuçları en yaygın olarak bir bağlam dizini olarak sunulur, bu da bu kavramı bu aşamada

tanımlamayı faydalı kılmaktadır. Bir bağlam dizini belirli bir kelime veya ifadenin derlemde bulunan örneklerinin bir listesini ve çevresindeki metin bağlamını içerir. Bağlam dizinleri artık bilgisayar tabanlı derlem araçları tarafından otomatik olarak üretilse de, çok daha uzun bir geçmişe sahiptir. Dijital teknolojinin ortaya çıkışından önce bağlam dizinleri manuel olarak derlenir ve referans eserler olarak yayımlanırdı. Erken örnekler arasında Hugh of St. Cher'in 13. yy'a ait Vulgata bağlam dizini (Rouse ve Rouse, 1979) gibi ortaçağ kutsal kitap bağlam dizinleri ile Shakespeare bağlam dizinleri ve Strong'un İncil bağlam dizini (Spevack, 1968) gibi klasik ve edebi metinlerin bağlam dizinleri yer alır.

Modern bağlam dizinleri, makine tarafından okunabilir metinlerden otomatik olarak üretilir ve bitmiş referans ürünleri olarak değil, öncelikle dinamik arama çıktıları olarak işlev yapar. Araştırma, akademik yazım, çeviri ve dilbilimsel analizde analitik araçlar olarak kullanılır; kullanıcıların kullanım kalıplarını ve bağlamı incelemesini sağlar. Bağlam dizinleri genellikle belirli sorgulara yanıt olarak üretildiğinden, kalıcı kaynaklar olarak depolanmak yerine geçici olarak derlem arayüzlerinde gösterilir.

Elektronik bağlam dizinleri, geleneksel basılı bağlam dizinlerine göre birçok avantaj sunar:

**1. daha hızlı ve esnek arama**

**2. arama ögesi etrafında genişletilebilir bağlamsal pencereler**

**3. çok daha büyük derlemlerin analizi**

Ayrıca, başlangıçta basılı olarak yayımlanan birçok bağlam dizini artık elektronik veya çevrimiçi olarak da erişilebilir hale gelmiş ve dilbilimsel araştırmalar için erişilebilirliklerini ve faydalarını daha da arttırmıştır.

### 3.1.2. KWIC Formatı (Bağlamda Anahtar Kelime Gösterimi)

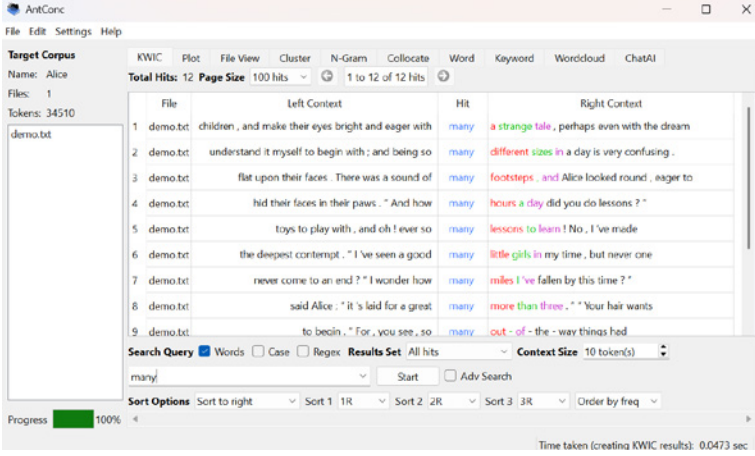
Derlem dilbiliminde erken dönem elektronik bağlam dizini araçları Bağlamda Anahtar Kelime (KWIC) formatına dayanırdı. Bu formatta, arama terimi genellikle yaklaşık 80 karakterden oluşan sabit genişlikli bir çizginin ortasında gösterilir. Bu tasarım, sınırlı ekran kapasitesine sahip erken dönem bilgisayar ekranlarının teknik sınırlamalarını yansıtıyordu. Çeviri odaklı derlem analizinde, KWIC bağlam dizinleri öncelikle yerel sözcük ortamlarını ve kaynak ile hedef metinlerdeki tekrarlayan kalıpları gözlemlemek için kullanılmıştır.

*KWIC formatının çeviri çalışmaları için avantajları:*

1. Tekrarlayan sözcük kalıplarının hızlı tanımlanmasını kolaylaştırır, bu özellik, birden fazla örnekte çeviri seçeneklerini incelemek için faydalıdır.
2. Çevirmenlerin ve araştırmacıların tekrarlayan eşdeğerleri karşılaştırmasına ve normalleşme, açıklama veya basitleştirme gibi eğilimleri belirlemesine olanak tanır.

*KWIC formatının çeviri odaklı analizdeki sınırlamaları:*

1. Sınırlı bağlamsal pencere, çeviri stratejilerini veya pragmatik anlamı tam olarak yorumlamak için genellikle yeterli bilgi sağlamaz.
2. Sözdiziminin yeniden yapılandırılması, cümle veya cümle düzeyindeki kaymaların ve daha uzun çeviri birimlerinin analizine uygun değildir. Bu durum, çeviri davranışı ve karar verme süreçlerini anlamada sıkça merkezi olan söylem düzeyindeki olguların incelenmesini sınırlar. (Mikhailov ve Cooper, 2016)



Şekil 18: AntConc programında 'many' kelimesini içeren Bağlamda Anahtar Kelime Dizini (KWIC) ekranı

### 3.1.3. Cümle Tabanlı Bağlam Dizimleri

Cümle tabanlı bağlam dizimleri, arama sonuçlarını kısaltılmış bağlam pencereleri yerine tam cümleler olarak sunar. Daha geniş bir metin eş metinsel ortama erişim sağlayan bu format araştırmacıların sözcük öğelerini tam sözdizim, anlamsal ve pragmatik bağlamlarında incelemelerine olanak tanır. Çeviri odaklı derlem analizinde, cümle tabanlı bağlam dizimleri özellikle tek tek bağımsız sözcük seçimlerinin ötesine geçip cümle veya cümle düzeyindeki çeviri birimlerine odaklanan analizler için çok uygundur.

Paralel derlemlerde, cümle tabanlı bağlam dizimleri, hizalanmış kaynak ve hedef metin segmentlerinin sistematik karşılaştırmasını kolaylaştırır. Bu, araştırmacıların çeviri kaymalarını, sözdizimsel düzenlemeleri ve bilgi yapısındaki değişiklikleri yalnızca anahtar kelime odaklı ekranlarla mümkün olandan daha büyük bir hassasiyetle gözlemlemesini sağlar.

Tam cümlelerin bulunabilirliği, açıklama, örtükleştirme ve modülasyon gibi çeviri olgularının analizini destekler; bu olaylar genellikle geniş bağlamsal bilgilere erişim gerektirir.

Karşılaştırma derlemlerinde, cümle tabanlı bağlam dizinleri, aynı dil içinde çevrilmiş ve çevirilmemiş metinler arasında doğal olarak bulunan dil kullanımının karşılaştırılmasına olanak tanır. Tam cümleleri inceleyerek, araştırmacılar çeviri dilin tipik tekrarlayan kalıplarını, standartlaştırma veya sadeleştirme eğilimleri dahil olmak üzere, nitel yorumu destekleyecek kadar bağlamsal derinliği koruyabilirler.

Cümle tabanlı bağlam dizinleri:

1. Anlam varyasyonunu ve kullanım kalıplarını yakalamak için geniş bağlamsal kanıtlara dayanan araştırmalar yapan sözcükbilimciler,

2. Paralel veya benzer derlemlerde sözdizim, anlamsal ve söylem düzeyindeki olguları analiz eden çeviribilimciler, tarafından tercih edilir. (Sinclair, 1991; Hunston, 2002; Baker, 1995)

BNCWeb, COBUILD ve COCA gibi derlem araçları, hem KWIC hem de cümle tabanlı bağlam dizini formatları sunar; bu da araştırmacıların derlem türüne ve araştırma hedeflerine en uygun bağlamsal detay seviyesini seçmesini sağlar.

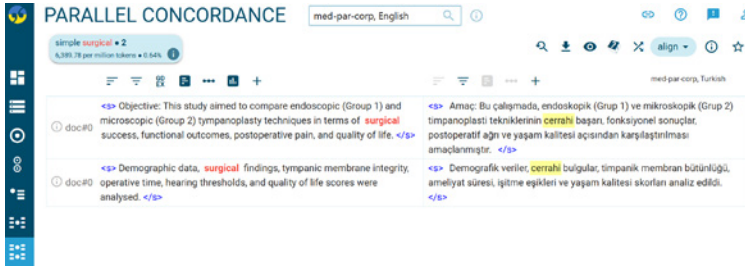
### **3.1.4. Paralel Bağlam Dizinleri**

Paralel bağlam dizini, kaynak metin bölümlerini karşılık gelen çevirileriyle birlikte sunmak üzere tasarlanmış paralel derlemlere dayanır. Bu format, kaynak ve hedef metinler arasındaki çeviri ilişkilerinin doğrudan incelenmesini sağlar ve çeviri kaymaları ile stratejilerinin sistematik analizini destekler. Paralel derlemler genellikle kaynak metin segmentlerini hizalanmış hedef metin karşılıklarıyla birlikte gösterir:

Paralel bağlam dizinlerinde yaygın olarak kullanılan çıktı formatları şunlardır:

1. Yan yana gösterim (paralel sütunlar), burada kaynak ve hedef metinler aynı anda gösterilir.
2. Ardışık gösterim, burada kaynak metnin ardından çevirisi yapılır.

TACT ve TamBiC gibi araçlar, KWIC gibi anahtar kelime merkezli formatlara kıyasla çeviri ilişkilerinin daha net bir temsilini sunarak tam cümle paralel bağlam dizinleri sağlar. Tam hizalanmış cümleler sunan bu araçlar genellikle kısaltılmış bağlam dizini gösterimlerinde gizlenen sözdizimsel düzenlemeler, anlamsal kaymalar ve çeviri stratejilerinin analizini kolaylaştırır (Hockey ve Martin, 1998; Küçük ve Can, 2008; Zanettin, 2012).



Şekil 19: SketchEngine programında Paralel Bağlam Dizini görünümü

### 3.1.5. Yazılım Araçları

Derlem dilbilimi ve çeviri çalışmalarında nicel yöntemlerin artan önemi, istatistiksel analizleri verimli şekilde yapabilen yazılımların varlığıyla yakından bağlantılıdır. Önceki araştırmalar manuel hesaplamalar ve istatistiksel tablolara danışma gerektirirken, günümüz araçları temel tablo programlarının bile çok çeşitli istatistiksel işlemleri gerçekleştirmesine olanak tanır. Elektronik tablo yazılımları küçük

ölçekli analizleri için yeterli olabilir, ancak genellikle büyük ve karmaşık veri setleri için uygun değildir; özellikle derlem verileri genellikle istatistiksel analiz yapılabilme için kapsamlı ön işlem gerektirir.

Derlemlerden elde edilen verileri, nicel analiz için uygun hale getirmek için sıklıkla yeniden düzenlenme, filtreleme, biçimlendirme veya dönüştürme yapılması gerekir. Teknik olarak bu tür işlemler elektronik tablo ortamlarında mümkün olsa da, genellikle zaman alıcı ve hata yapmaya açıktır. Masaüstü veritabanı yazılımları, büyük veri kümelerini yönetmek ve sorgu tabanlı işlemlerle yapılandırılmış veri işleme yapmak için daha fazla esneklik sunar, ancak entegre istatistiksel işlevselliğin eksikliği nedeniyle derlem araştırmalarında kullanımları yaygınlaşmamıştır.

Bu nedenle, hem veri işleme hem de analiz için özel olarak tasarlanmış istatistiksel yazılım paketleri genellikle en etkili çözüm olarak kabul edilir. IBM SPSS Statistics (IBM Corp., 2020) ve R (R Core Team, 2023) gibi programlar, araştırmacıların büyük veri setlerini yönetmelerini, rutin veri işleme görevlerini otomatikleştirmelerini ve istatistiksel testleri nispeten kolayca gerçekleştirmelerini sağlar. Her iki platform da geniş çapta desteklenir, platformlar arası uyumludur ve sonuçların görsel temsillerini üretebilir ve yorumlamaya katkı sağlar.

Kullanılan yazılım ne olursa olsun, derlem verilerinin güvenilirliği önemini korumaktadır. Derlem tabanlı araştırmada, özellikle büyük veya heterojen veri setleriyle çalışırken derlem bileşiminin dikkatli değerlendirilmesi gerekir. Birçok durumda, belirli bir zaman dilimine, metin türüne, dil çeşitliliği veya yazara odaklanmak gibi belirli araştırma hedeflerine uygun bir alt derlem tanımlamak metodolojik olarak tercih edilebilir. İyi tasarlanmış, hazır derlemler kullanıldığında

bile, arařtırmacılar tüm alt bileřenlerin içsel dengeli veya ho-mojen olmasına dikkat etmelidir.

Bu nedenle derlem büyüklüğü ve yapısının deęerlendi-rilmesi temel bir ön adımdır. Sürekli tekrarlanan kelimeler, cümleler veya karakterler gibi ölçütler genellikle derlem bo-yutunu tanımlamak için kullanılır ve birçok derlem arayüzü bu istatistikleri hesaplamak için yerleşik araçlar sağlar. Metin-ler elektronik formatta mevcut olduğunda, standart derlem analiz yazılımını temel nicel bilgi üretmek için kullanılabilir. Bu tür kontroller, sonraki nicel bulguların yeterince büyük, kar-şılaştırılabilir ve metodolojik olarak sağlam veri setlerine da-yanmasını sağlamak için çok önemlidir.

# BÖLÜM IV - Çeviribilim Araştırmalarında Derlem Kullanımı

*“Corpora do not replace theory; they make theory accountable.”*

John Sinclair

## 4.1. Derlem Verilerinde Merkezi Eğilimlerin Araştırılması

Nicel derlem tabanlı araştırma, verilerdeki genel eğilimlerin incelenmesiyle başlamalıdır; çünkü bu, veri yeterliliğinin ilk değerlendirmesini sağlar ve daha fazla istatistiksel test gerektiren alanların belirlenmesine yardımcı olur. Merkezi eğilim ölçümleri bu aşamada özellikle faydalıdır; çünkü değerlerin bir veri seti içinde nasıl dağıldığına genel bir bakış sunar. En yaygın kullanılan üç ölçü ortalama, medyan ve moddur.

Tüm değerlerin aritmetik ortalaması olarak hesaplanan ortalama, verilerde merkezi bir noktayı temsil eder ancak aşırı değerlere duyarlıdır. Metin uzunlukları veya çok verimli yazarlar gibi önemli değişkenlikle karakterize edilen veri setlerinde, ortalama tipik davranışın yanıltıcı bir resmini sunabilir. Sıralı veri setindeki orta değer olarak tanımlanan medyan, aykırı değerlerden daha az etkilenir ve dolayısıyla heterojen derlem verilerine daha uygundur. En sık görülen değeri belirleyen mod, baskın desenleri tespit etmek için özellikle faydalıdır, ancak veri setleri birden fazla mod gösterebilir.

Merkezi eğilimlerin yanı sıra, değerlerin genel dağılımı, minimum ve maksimumları dahil olmak üzere, ayrıca gözlemlerin merkezi bir nokta etrafında ne kadar kümелendiğinin de dikkate alınması önemlidir. Cümle uzunluğu veya

kelime kullanım sıklığı gibi birçok dilsel değişken, ortalama, medyan ve modun birleştiği normal bir dağılım gösterme eğilimindedir. Normallikten sapmalar, dikkatli yorum gerekti- ren çarpık veya eşit olmayan dağılımları gösterebilir.

Veri davranışını daha kapsamlı anlamak için araştırmacılar genellikle merkezi eğilim ölçümlerini ek tanımlayıcı istatistiklerle tamamlar. Bunlar arasında, veri içindeki dağılım derecesini gösteren standart sapma ile dağılımın simetrisini ve zirvesini tanımlayan çarpıklık ve basıklık ölçümleri bulunur. Derlem tabanlı çalışmalarda yaygın olarak kullanılan diğer göstergeler arasında varyans, z-puanları ve tip-token oranı (TTR) bulunur. Bu ölçümler genel eğilimleri belirlemek ve derlem tabanlı çeviri araştırmalarında sonraki analizleri yönlendirmek için sağlam bir istatistiksel temel sağlar.

#### **4.2. Derlem Tabanlı Araştırmada Anahtar Kelime Analizi**

Anahtar kelime analizi, bir araştırma derlemindeki kullanım sıklığı bir referans derlemindeki sıklıktan önemli ölçüde farklı olan sözcük öğeleri tanımlayan ve derlem dilbiliminde yaygın kullanılan bir istatistiksel tekniktir. Birincil amacı, bir metin bütününe diğerinden ayıran özellikleri vurgulamaktır. WordSmith (Scott, 2008) ve AntConc (Anthony, 2019) gibi derlem analiz araçları, bu prosedürü kullanım sıklığı listelerini karşılaştırarak ve gözlemlenen farkların istatistiksel anlamlılığını hesaplayarak uygular; genellikle ki-kare veya log-olasılık testleri kullanılır. Aşırı temsil edilen öğeleri (anahtar kelimeleri) tanımlamanın yanı sıra, bu araçlar genellikle negatif anahtar kelimeler olarak adlandırılan az temsil edilen öğeleri de tespit edebilir.

KEYWORDS med-par-corp. English

SINGLE-WORDS ✓ MULTI-WORD TERMS ✓

reference corpus: English Web 2021 (enTenTen21)

Lemma	Lemma	Lemma	Lemma	Lemma
1 endoscopic ***	11 preoperatively ***	21 membrane ***	31 ac ***	41 media ***
2 comq-12 ***	12 tympanic ***	22 db ***	32 p ***	42 differ ***
3 tympanoplasty ***	13 conduction ***	23 demographic ***	33 bc ***	43 functional ***
4 postoperative ***	14 retrospectively ***	24 surgical ***	34 outcome ***	44 pain ***
5 abg ***	15 microscopic ***	25 threshold ***	35 surgery ***	45 yield ***
6 otitis ***	16 perforation ***	26 analog ***	36 inventory ***	46 bone ***
7 combi ***	17 forty-five ***	27 integrity ***	37 evaluate ***	47 comfort ***
8 yas ***	18 anatomical ***	28 chronic ***	38 significantly ***	48 score ***
9 questionnaire-12 ***	19 auditory ***	29 fifteen ***	39 hearing ***	49 effectively ***
10 air-bone ***	20 operative ***	30 analyse ***	40 undergo ***	50 technique ***

Rows per page: 50 1-50 of 100

Şekil 20: SketchEngine Programında Anahtar Kelimesi Arama Sonucu

Anahtar kelime analizinin etkililiği, araştırma derlemi ile referans derlemi arasındaki ilişkiye çok bağlıdır. İlk genellekle dar bir şekilde tanımlanır ve belirli araştırma amaçlarıyla uyumlu olurken, ikincisi genellikle daha genel dil kullanımı veya zıt bir veri setini temsil eder. Çeviri çalışmalarında, anahtar kelime analizi farklı çeviri yönleri, yazarlar, türler ve kurumsal bağlamlar dahil olmak üzere çeşitli karşılaştırmalı senaryolara uygulanmıştır. Ayrıca, çeviri metinlerde ideolojik olarak belirgin sözcük tercihlerini belirlemede de faydalı olmuştur.

Anahtar kelime listeleri, kelime biçimi (tür) sıklık listelerinden veya kök biçim (lema) tanımlı listelerden oluşturulabilir ve her biri farklı analitik içgörüler sağlar. Kelime biçimlerine dayalı analizler, dilbilgisi kalıplarını, metin türlerine özgü yapıları ve sabit ifadeleri ön plana çıkarırken, kök biçim tanımlı analizler özel varyasyondan uzaklaşır ve metinlerin tematik veya konusal özelliklerini yansıtmaya olasıdır ve daha yüksektir. Sözcük öğelerinin ötesinde, anahtar kelime analizi dilbilgisel, anlamsal veya açıklama tabanlı kategorilere de genişletilebilir ve bu da metodolojik esnekliğini vurgular.

Faydasına rağmen, anahtar kelime analizinin belirgin sınırlamaları vardır. Örneğin, sınırlı verilerin önemsiz veya bilgilendirici olmayan anahtar kelimelere yol açabileceği küçük derlemlere uygulandığında daha az güvenilirdir. Ayrıca, güvenilir sonuçlar alınması için genellikle araştırma derleminde çok daha büyük bir referans derlemi gerekir. Son olarak, uygun bir referans derleminin seçimi her zaman kolay değildir ve analiz sonuçlarını etkileyebilir. Bu hususlar, derlem tabanlı çeviri araştırmalarında anahtar kelime analizi kullanılırken dikkatli derlem tasarımı ve metodolojik şeffaflığın önemini vurgular.

### **4.3. Derlem Tabanlı Araştırmada Eşdizimlilik Analizi**

Bağlam dizini verilerinin nicel analizi, gözlemlenen çeviri kalıplarının sistematik eğilimlerin mi yoksa sadece bireysel çevirmen tercihlerinin sonucu mu olduğunu değerlendirmede kritik bir rol oynar. Derlem tabanlı çeviri çalışmalarında, bağlam dizini analizinden elde edilen niteliksel gözlemler istatistiksel testlerle güçlendirilebilir. Bu amaçla yaygın olarak kullanılan yöntemlerden biri, gözlemlenen frekans dağılımlarındaki farkların istatistiksel olarak anlamlı mı yoksa tesadüfi mi olduğunu değerlendiren ki-kare testidir.

Ki-kare testleri, gözlemlenen kullanım sıklıklarının bir çapraz tablosundan elde edilen beklenen değerlerle karşılaştırılması yoluyla kategorik verilerin bağımsızlığını incelemek için kullanılır. Çeviri verilerine uygulandığında, bu prosedür, çevirmenler veya alt derlemler arasında çeviri eşdeğerlerinin kullanımındaki farklılıkların istatistiksel olarak güvenilir olup olmadığını belirlemeyi mümkün kılar. Düşük olasılık değerleri (p-değerleri), gözlemlenen dağılımın rastgele olma olasılığının düşük olduğunu gösterir ve bu nedenle sistematik varyasyonun yorumlanmasını destekler. Çağdaş derlem

arařtırmalarında, bu tür testler rutin olarak istatistiksel yazılım kullanılarak yapılır ve arařtırmaclar manuel hesaplama yerine yorumlamaya odaklanabilirler.

Çeviri eşdeğerlerinin analizinin ötesinde, nicel yöntemler de eşdizimli kelimelerin incelenmesinde merkezi bir yere sahiptir. Ortak konum, genel olarak sözcük öğelerinin herhangi bir birlikte ortaya çıkması olarak tanımlanabilirken, derlem tabanlı yaklaşımlar giderek daha fazla istatistiksel ölçütlere dayanarak tesadüfen beklenenden daha güçlü ortak oluşumları tespit eder. Bu çerçevede, eşdizimlilik, karşılıklı bilgi, t-puanı, z-puanı ve log-olasılığı gibi çeşitli istatistiksel ilişki ölçütleri kullanılarak hesaplanan eşdizimlilik güçlerine göre tanımlanır.

The screenshot shows the SketchEngine Word Sketch interface. At the top, the title is "WORD SKETCH" and the search term is "med-par-corp, English". Below the title, there is a search bar containing "surgery as noun 2x". The main area displays three panels of word sketches:

- modifiers of "surgery"**: microscopic, microscopic surgery
- verbs with "surgery" as object**: have, had microscopic surgery
- prepositional phrases**: ... after "surgery"

Şekil 21: SketchEngine programında 'surgery' kelimesinin eşdizimlilik analiz ekranı

Bu ölçütler arasında, Dunning'in log-olasılık (referans) istatistiği büyük veri kümelerindeki kullanım sıklığı farklılıklarına duyarlılığı nedeniyle derlem dilbilimi ve çeviri çalışmalarında yaygın olarak kullanılmaktadır (Dunning, 1993). Log-olasılık ölçüsü, bir kelimenin gözlemlenen ortak oluşma sıklıklarını ve potansiyel eşdizimliliğini genel derlem dağılımına göre beklenen kullanım sıklıklarıyla karşılaştırır. Sonuç olarak, log-olasılık, derlem verilerindeki istatistiksel olarak

anlamalı eşdizimli kelime desenlerini belirlemede sağlam bir yöntem sağlar.

Ki-kare testi ve istatistiksel eşdizimlilik analizi, nicel yöntemlerin bağlam dizini temelli bulguların güvenilirliğini nasıl artırabileceğini göstermektedir. Niteliksel yorumu istatistiksel doğrulama ile birleştiren derlem tabanlı çeviri çalışmaları, gerçek çeviri kalıplarını özgün veya tesadüfi varyasyonlardan ayırt etmek için kullanılabilir.

#### **4.4. Sözcükbilim Analizlerinde Paralel Derlem Kullanımı**

İki dilli ve çok dilli derlemler öncelikle araştırma amaçlı tasarlanmış olsa da, aynı zamanda iki dilli sözlüklerle karşılaştırılabilir referans aracı olarak da işlev görebilirler. Paralel derlemler, kullanıcıların belirli kelimelerin veya ifadelerin çeviri eşdeğerlerini aramasına olanak tanır, ancak sözlüklerin aksine, geniş bağlamsal örnekler ve alternatif çeviriler sunarlar. Tam cümlelere gömülü çoklu çeviri çözümleri sunan paralel derlemler çeviri ve yabancı dil yazımda daha bilinçli karar almayı destekler.

Paralel derlemlerin geleneksel sözlüklere göre önemli avantajlarından biri, çeviri pratiğindeki farklılıkları ortaya koyabilme yetenekleridir. Derlem verileri sınırlı bir standart eşdeğer seti sunmak yerine, çeviri seçeneklerinin bağlam, tür ve söylem fonksiyonuna nasıl bağlı olduğunu gösterir. Bu durum, paralel derlemleri özellikle deyimsel kullanımları tanımlamak ve bazı bağlamlarda uygun olmayabilecek mekanik olarak uygulanan sözlük eşdeğerlerinden kaçınmak için değerli kılar.

Ancak, paralel derlemlerin arama aracı olarak etkinliği büyük ölçüde derlem boyutu ve bileşimine bağlıdır. Küçük derlemler genellikle nispeten sık kullanılan kelimeler ve

ifadelerle sınırlıdır, bu da daha az yaygın veya çok özel kullanımlar için faydalarını sınırlar. Ayrıca, paralel derlemlerin güvenilirliği de temkinli bir şekilde ele alınmalıdır. Derlemler çevrilmiş metinlerden oluştuğu için, kaçınılmaz olarak bireysel çevirmenlerin kararlarını ve olası hatalarını yansıtır. Sonuç olarak, derlem kanıtları, özellikle kültüre bağlı veya tarihsel olarak değişen terimler söz konusu olduğunda modası geçmiş, yanıltıcı veya bağlamsal olarak uygunsuz eşdeğerler verebilir.

Paralel derlemler, özellikle doğrulanmış kaynak-hedef segmentlerini depolayan ve benzer girdiler için önerilen çeviriler sunan çeviri belleği (TM) sistemleri olmak üzere, bilgisayar destekli çeviri araçlarıyla birlikte giderek daha fazla kullanılmaktadır. TM'ler, özel paralel derlemlerle yapısal benzerlikleri paylaşırsa da, benzer sınırlamalara sahiptir. TM önerileri daha önce çevrilmiş segmentlere dayandığı için, hatalar veya optimal olmayan çözümler kullanıcı tarafından eleştirel olarak değerlendirilmedikçe tekrar kullanılabilir (Bowker, 2002; Somers, 2003; O'Brien, 2012).

Genel olarak, yeterince büyük ve iyi tasarlanmış paralel derlemler değerli referans kaynakları olarak hizmet verebilirken, yarı-sözlük olarak kullanımları dikkatli yorumlama ve eleştirel farkındalık gerektirir. Birincil işlevleri, belirleyici rehberlik yerine ampirik araştırma olarak kalmaktadır. Derlem tabanlı çeviri çalışmalarında gösterildiği gibi, paralel derlemler özellikle sözcüksel, morfolojik, sözdizimsel ve pragmatik olguları ile çeviri teorisinin merkezinde yer alan konuları incelemek için uygundur.

#### 4.5. Çokdilli Derlemler ve Terminoloji Çalışmaları

Terminoloji çalışmaları, belirli bir alan içindeki uzmanlaşmış terimlerin tanımlanması, ve sistematikleştirilmesine odaklanır. Sözlükle yakından ilişkili olmasına rağmen, terminoloji araştırması kavramsal sistemlere ve uzman doğrulamasına verdiği önemle ayırt edilir (Cabr , 1999; Sager, 1990). Terminolojik çalışmanın temel amacı, belirli alanlar i in kapsamlı ve tutarlı terim setleri derlemektir; bu s re  kavramsal analiz, alan uzmanlarıyla istişare ve yapılandırılmış isimlendirmele-  
rin oluřturulmasını i erir. Aynı zamanda, terminoloji doęası gereęi dilseldir;  nk  terimler, genel kelime daęarcıęıyla aynı anlamsal, dilbilgisel ve pragmatik kısıtlamalara tabi olan s zc k birimleridir (Bowker, 2015).

Dilbilimsel a ıdan, terminolojik arařtırmalar eřseslilik,  oklu anlam ve eřanlamlılık gibi konuları ele almalıdır; bunların t m  derlemlerde terimlerin tanımlanması ve yorumlanmasını zorlařtırır. Eřanlamlılık genellikle s zdizimi analiziyle c z lebilse de,  oklu anlamlara sahip olan terimler daha  nemli bir zorluk teřkil eder;  nk  bir ok teknik terim g nl k dilden t remiř daha genel anlamlara da sahiptir. Eřanlamlılık, bazen terminolojik sistemlerde hassasiyeti azaltma potansiyeli nedeniyle teřvik edilmese de, alanlar arasında ve i inde g r l r ve bu da dikkatli dilsel analizin gereklilięini daha da vurgular.

 ok dilli derlemlerin terminolojik arařtırma i in a ık potansiyeline raęmen, bu alandaki kullanımları sınırlı kalmaktadır. Terminoloji arařtırmacıları geleneksel olarak uzman metinlerine, s zl klere, ansiklopedilere ve uzman danıřmanlıęına dayanmıř, bazen tek dilli derlemlerden destek almıřlardır. B y k  l ekli derlem tabanlı terminoloji  alıřmaları h l  b y k  l c de pratik ve metodolojik kısıtlamalar nedeniyle nadirdir. B y k engellerden biri, genellikle mevcut olmayan

ve sıfırdan oluşturulması gereken uygun paralel veya karşılaştırılabilir uzman metin derlemelerini derlemek için gereken kapsamlı hazırlık çalışmalarıdır. Bu zorluk, birçok alanda terminolojinin hızlı evrimiyle daha da karmaşık hale gelir ve bu da sık sık derlem güncellemelerini gerektirir.

Standart derlem araçları, terminolojik açıklamanın merkezinde yer alan çok kelimeli terimlerden ziyade tek kelimeli birimleri tanımlamak için daha uygundur. İstatistiksel yöntemler ve eşdizimlilik belirleme yöntemleri çok kelimeli terimlerin çıkarılmasına yardımcı olabilse de, etkililikleri düşük kullanım sıklıkları ve morfolojik olarak zengin dillerin karmaşıklığı nedeniyle sınırlıdır. Ayrıca, paralel derlemelerden otomatik olarak çıkarılan çeviri eşdeğerlerinin güvenilirliği düşük olabilir. Çeviri metinler, özellikle uzman olmayanlar tarafından üretildiğinde, yanlışlıklar veya standart olmayan terimler içerebilirken, yalnızca paralel derlemelerden türetilen terminolojik kaynaklar hataları ve eksiklikler içermeye risk taşır.

Bu sınırlamaları azaltmak için birleşik metodolojik yaklaşım önerilir. Derlem tabanlı terminolojik araştırmalar, hem paralel hem de karşılaştırılabilir derlemelerin kullanımından faydalanır; çevrilmiş veriler alan uzmanları tarafından üretilen orijinal metinlerle sistematik olarak doğrulanır. Kullanım sıklığı listeleri, aday madde başı kelimeleri belirlemek için kullanılabilir; bu kelimeler genellikle çok kelimeli terimler için çekirdek olarak kullanılır. Bu madde başı kelimelerin bağlam dizini analizi daha sonra terminolojik kümelerin tanımlanmasını sağlar. Ortaya çıkan terimler, yapılandırılmış veritabanlarına girilmeden önce tek dilli derlemeler ve yerleşik terminolojik kaynaklar kullanılarak doğrulanmalıdır. Böyle bir yaklaşım, derlem tabanlı yöntemlerin verimliliği ile

terminolojik araştırma için gereken güvenilirlik arasında denge sağlar.

#### **4.6. Paralel Derlem Araştırmalarında Morfoloji ve Sözdizimi**

Dilbilgisel tanım geleneksel olarak kelime yapısıyla ilgili morfoloji ve cümle ile cümle kalıplarını inceleyen sözdizimi olarak ikiye ayrılmıştır (Matthews, 1991; Huddleston ve Pullum, 2002). Derlem tabanlı araştırmalarda, morfolojik olgular genellikle sistematik incelenmeye daha uygundur; çünkü morfolojik desenler genellikle net tanımlanır ve nispeten kolay tanımlanır. Buna karşılık, sözdizimi desenleri daha karmaşık ve daha az doğrudan alınabilir olma eğilimindedir; güvenilir derlem aramaları için iyi tanımlanmış biçimsel işaretleyiciler gerektirir (Biber vd., 1998; McEnery & Hardie, 2012).

Paralel derlem, diller arasındaki morfolojik desenlerin karşılaştırmalı incelenmesi için değerli kaynaklar sunar. Bir dilde açıkça işaretlenen morfolojik kategoriler, başka bir dilde fonksiyonel eşdeğerleriyle sistematik olarak karşılaştırılabilir. Bu tür analizler, dillerin benzer anlamları farklı biçimsel yöntemlerle kodlayabileceğini gösteren beklenmedik karşılıkları ortaya çıkarabilir. Bu bulgular, paralel derlem morfolojik ifadedeki diller arası düzenlilikleri ve farklılıkları tanımlamada faydasını vurgulamaktadır.

Paralel derlem kullanılarak yapılan sözdizimi araştırmaları genellikle cümle düzeyinde karşılıklara odaklanır ve arama kriterleri olarak uygulanabilecek net yapısal sinyallerin belirlenmesini gerektirir. Kaynak ve hedef cümlelerinin hizalanmasını inceleyerek, araştırmacılar zaman, kip veya modalite gibi dilbilgisel kategorilerin diller arasında nasıl farklı şekilde gerçekleştirildiğini keşfedebilirler. Paralel

derlem analizi, bir dilin morfolojik işaretleme kullanırken diğerinin sözdizimi yapılarına dayandığı durumlar da dahil olmak üzere, sözdizimi seçiminde sistematik zıtlıkları ortaya çıkarabilir.

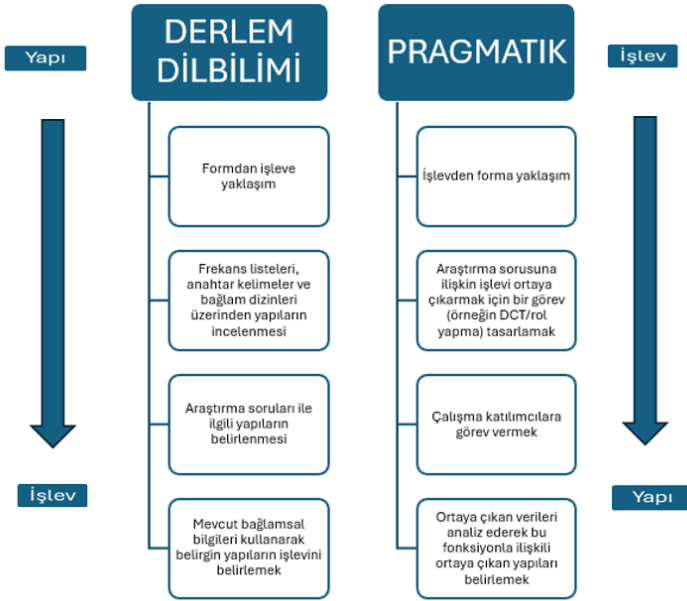
Morfoloji ile sözdizimi arasındaki ayırım dillerarası araştırmalarda daha az belirgin hale gelir. Bir dilde morfolojik olarak kodlanan özellikler, başka bir dilde sözdizimi yapılarına karşılık gelebilir ve tam tersi de geçerlidir. Bu örtüşme, çok dilli veri analiz ederken katı kategorik sınırların korunmasına yönelik girişimleri zorlaştırır. Sonuç olarak, karşılaştırmalı derlem çalışmaları genellikle fonksiyonel eşdeğerliği resmi sınıflandırmadan üstün tutan esnek bir analitik bakış açısı benimser.

Genel olarak, paralel derlemler, diller arasındaki dilbilgisel zıtlıkları araştırmak için güçlü bir ampirik temel sağlar. Uyumlu dilsel yapıların sistematik karşılaştırmasını mümkün kılarak, morfolojik ve sözdizimsel anlamların diller arasında nasıl dağıldığına dair incelikli analizleri destekler ve böylece hem betimleyici dilbilime hem de çeviri çalışmalarına katkıda bulunurlar.

#### **4.7. Edimbilim ve Çok Dilli Derlem Analizi**

Edimbilim (pragmatik), konuşmacıların ve yazarların kelimeleri dilsel içeriğin ötesinde anlamı nasıl aktardıklarına odaklanarak, dili iletişimsel bağlamlarda bir sosyal eylem biçimi olarak inceler. İşbirliği ve yanlış anlama, nezaket ve sosyal gelenekler, konuşma eylemleri, tartışma ve retorik stratejiler gibi olgularla ilgilendirir. Edimbilim, konuşma eylemi teorisi, konuşma analizi, sosyodilbilim ve retorik gibi disiplinlerle yakından bağlantılıdır.

Edimbilim olguları, derlem tabanlı analiz için özel zorluklar oluşturur; çünkü edimbilimsel anlam genellikle dolaylı olarak ifade edilir ve tonlama, jest ve paylaşılan bağlamsal bilgi gibi dil dışı ipuçlarına dayanabilir. Derlemler, konuşma etkileşimini incelemek için kullanılabilir de, uygun verilerin bulunabilirliği sınırlıdır. Kurgusal diyalog, kolayca erişilebilir olsa da, otantik konuşmadan ziyade stilize taklitleri temsil eder ve bu nedenle edimbilimsel araştırmalar için sınırlı güvenilirliğe sahiptir. Yazıya dökülmüş sözlü etkileşim daha gerçek veri sunar ancak yazıya dökmenin karmaşıklığı ve konuşmanın doğası gereği değişkenlik nedeniyle yeterince büyük ölçekte derlenmesi zordur.



Şekil 22: Söylem Pragmatisinde Yaklaşımlar (Mikhailov ve Cooper, 2016).

Çok dilli ve çeviri odaklı derlem çalışmalarında ek metodolojik endişeler ortaya çıkar. Çevrilmiş diyaloglar, hedef dilin yerine kaynak dilin edimbilimsel normlarını yansıtabilir; bu da hedef bağlamda edimbilimsel olarak doğal olmayan ifadelere yol açabilir. Bu tür müdahaleler, çevrilmiş diyaloga temkinli yaklaşmanın ve mümkün olduğunda onu çevrilmiş söylemle karşılaştırmanın gerekliliğini vurgular. Yorumlanan etkileşim, özellikle çok dilli tartışmalarda, özellikle zengin bir edimbilimsel veri kaynağı oluşturur; çünkü yorumcular sürekli olarak anlamsal doğruluk ile edimbilimsel uygunluk arasında denge kurmak zorundadır. Ancak, büyük ölçekli yorumlanmış verilere erişimin pratik zorluğu, bunun derlem araştırmalarında kullanımını sınırlar.

Bu kısıtlamalara rağmen, bazı edimbilimsel olgular derlem tabanlı araştırma için oldukça uygundur. Edimbilimsel parçacıklar gibi net biçimsel sinyallere sahip dile özgü işaretleyiciler sistematik olarak analiz edilebilir. Araştırmacılar bu tür işaretleyicileri aramada referans noktası olarak kullanarak, edimbilimsel anlamın diller arasında nasıl kodlanıp yeniden ifade edildiğini inceleyebilir ve çeviride kullanılan telafi stratejilerini ortaya çıkarabilirler. Çok dilli derlemler, edimbilimsel varyasyon ve dillerarası edimbilimsel aktarım konusunda kısmi olsa da değerli, içgörüler sağlar.

## SONUÇ

*“The future belongs to those who can connect evidence  
with imagination.”*

C. Wright Mills

### **Derlem Tabanlı Çeviri Çalışmalarının Geleceğine İlişkin Değerlendirmeler**

Çeviride derlem tabanlı çeviri çalışmalarının geleceği birçok araştırmacı tarafından ayrıntılı biçimde ele alınmış; bu tartışmalar ışığında derlemlerin kapsamının genişletilmesi, yöntemsel yeniliklerin benimsenmesi, dijital çeviri ortamlarının dikkate alınması ve daha kapsayıcı bir dil politikası izlenmesi gerektiği yönünde çeşitli saptamalar yapılmıştır (Baker, 1995; Laviosa, 2002; Zanettin, 2012; Gaspari, 2022; Granger & Lefer, 2022).

Çeviri ve derlem tabanlı çeviri çalışmalarının (CBTS) gelecekteki gelişimi birçok belirsizlik içerse de, alan teorik, metodolojik ve uygulamalı ilerleme için önemli fırsatlar sunmaya devam etmektedir. Çözülmemiş zorlukların ele alınması ve mevcut boşlukların kapatılması, CBTS'nin çeviri uygulamasının gelişen doğasına yanıt vermeye devam edebilmesi için hayati öneme sahiptir.

Önemli bir risk, CBTS'nin dijital ve çevrimiçi ortamlarda çevirinin hızlı yayılmasına ayak uyduramamasıdır. Çeviri giderek yeni biçimler ve işlevler kazandıkça, özellikle teknoloji aracılı iletişimde, alanın yalnızca yerleşik yöntemlere ve tanıdık veri tiplerine güvenmekten kaçınması gerekir. Sürekli önemlilik, karmaşık ve daha az yapılandırılmış verilerle

ilgilenme ve yeni ve ortaya çıkan çeviri senaryolarına uygun analitik yaklaşımlar geliştirme istekli olma isteğine bağlıdır.

Gelecekteki arařtırmalar için önemli bir öncelik, CBBDT'nin dilsel kapsamını genişletmektir. Bugüne kadar arařtırmalar, İngilizce ve diđer yaygın kullanılan küçük bir grup dil tarafından domine edilmiştir; bu da büyük ölçüde derlemlerin erişilebilirliđi, finansman yapıları ve akademik yayıncılık normları sayesinde gerçekleşmiştir. Bu kısıtlamalar anlaşılır olsa da, azınlık ve az kaynaklı dillerin sistematik olarak dışlanmasına katkıda bulunmuştur. Çevirinin giderek dijitalleşmesi, çok dilli verilere erişim için yeni olanaklar sunar ve daha kapsayıcı, etik açıdan bilinçli bir arařtırma gündeminin geliştirilmesini destekler.

Aynı zamanda, çeviri evrenselleri, açıklama, yönlendirme ve arabuluculuk gibi temel arařtırma temalarına devam etmek hayati öneme devam etmektedir. Bu alanlardaki ilerleme, ilgili arařtırma dalları arasında daha fazla metodolojik yakınsamadan ve artan diyalogdan fayda sağlayacaktır. Uzun süredir devam eden teorik ve operasyonel sorunların ele alınması, sahanın devam eden canlılığını sağlamak için merkezi olacaktır.

Sonuç olarak, derlem tabanlı çeviri çalışmalarının geleceđi, yenilik, kapsayıcılık ve metodolojik esnekliđe kolektif bir bağlılıđa bağlıdır. Yeni çeviri biçimleriyle ilgilenerek, dilsel kapsamı genişleterek ve analitik araçları geliştirerek, BDTS kapsamını ve alakasını genişleterek hem çeviri çalışmalarına hem de kullanılan dilin daha geniş incelenmesine anlamlı katkı sağlayabilir.

## KAYNAKÇA

- Anthony, L. (2019). AntConc (Version 3.5.8) [Bilgisayar yazılımı]. Waseda University.
- Algeo, J. (1998). A corpus of English from 1700 to the present. S. Johansson & H. Hasselgård (Ed.), *Corpora and cross-linguistic research* içinde (ss. 1–14). Rodopi.
- Altenberg, B., & Granger, S. (Ed.). (2002). *Lexis in Contrast: Corpus-Based Approaches* (Studies in Corpus Linguistics, Vol. 7). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Baker, M., Francis, G., & Tognini-Bonelli, E. (Ed.). (1993). Text and technology: In honour of John Sinclair (pp. 233–250). Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Baker, M. (1995). Corpora in Translation Studies: An overview and some suggestions for future research. *Target*, 7(2).
- Baker, M. (1999). The role of corpora in investigating the linguistic behaviour of professional translators. *International Journal of Corpus Linguistics*, 4(2), 281–298.
- Baker, M. (1993). Corpus linguistics and translation studies. G. Francis & E. Tognini-Bonelli (Ed.), Text and technology: In honour of John Sinclair içinde (ss. 233–250). Amsterdam: John Benjamins
- Barlow, M. (2003). *ParaConc: A concordancer for parallel texts* [Bilgisayar yazılımı]. Houston: Athelstan.
- Bendazzoli, C. (2010). Corpora and interpreting research: The EPIC project. C. Bendazzoli, N. Russo & G. Sandrelli (Ed.), *Methodology and data analysis in interpreting studies* içinde (ss. 39–63). Trieste: EUT – Edizioni Università di Trieste.
- Benko, V. (2014). Aranea: Yet another family of (comparable) web corpora. *Text, Speech, and Dialogue* içinde (ss. 247–256). Springer.

- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Biber, D., Conrad, S., & Reppen, R. (2000). *Longman grammar of spoken and written English*. London, UK: Longman.
- Biber, D., Conrad, S., & Reppen, R. (2006). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Borin, L. (2002). *Parallel corpora, parallel worlds*. Amsterdam: Rodopi. Cambridge University Press
- Borin, L., Forsberg, M., & Roxendal, J. (2012). Korp – The corpus infrastructure of Språkbanken. Proceedings of the Eighth International Conference on Language
- Bowker, L. (2002). *Computer-aided translation technology: A practical introduction*. Ottawa: University of Ottawa Press.
- Bowker, L., & Pearson, J. (2002). *Working with specialized language: A practical guide to using corpora*. London: Routledge.
- Bowker, L. (2015). Terminology and lexicography. H. J. Kockaert & F. Steurs (Ed.), *Handbook of terminology* içinde (pp. 1–22). Amsterdam: John Benjamins.
- Brill, E. (1995). Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging. *Computational Linguistics*, 21(4), 543–565.
- Burnard, L. (2000). Reference guide for the British National Corpus (World Edition). Oxford University Computing Services.
- Burnard, L., & McEnery, T. (2000). SARA and the British National Corpus. *Literary and Linguistic Computing*, 15(1), 7–16.
- Burnard, L., & Sperberg-McQueen, C. M. (2002). *TEI Lite: An introduction to text encoding for interchange*. Oxford: Text Encoding Initiative Consortium.
- Burnard, L., & Bauman, S. (Ed.). (2014). *TEI P5: Guidelines for electronic text encoding and interchange*. Text Encoding Initiative Consortium.

- Cabré, M. T. (1999). *Terminology: Theory, methods and applications*. Amsterdam: John Benjamins.
- Carroll, J. B., Davies, P., & Richman, B. (1971). *The American Heritage word frequency book*. Boston, MA: Houghton Mifflin.
- Čermák, F. (2007). Czech National Corpus: A synopsis of the history and present state. Nakladatelství Lidové noviny. Resources and Evaluation (LREC 2012) içinde (ss. 474–478). Istanbul: ELRA.
- Chomsky, N. (1968). *Syntactic Structures*. The Hague Mouton and Co.
- Christ, O., Schulze, B. M., Hofmann, A., & Koenig, E. (1999). The IMS Corpus Workbench: Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, 33A(4), 497–505.
- Corpus query processor (CQP). Proceedings of the Eighth EACL Conference.
- Davies, M. (2004). BYU-BNC: The British National Corpus on the Web. *International Journal of Corpus Linguistics*, 9(2).
- Davies, M. (2008). The Corpus of Contemporary American English (COCA): 560 million words, 1990–present. *International Journal of Corpus Linguistics*, 14(2), 159–190.
- Dayter, D. (2016). Discursive practices in the United Nations: A corpus-based approach. *Journal of Language and Politics*, 15(3), 287–310.
- Dembry, C., & Love, R. (2015). Size matters: Spoken interpreter corpora and reference corpora. *Corpora*, 10(1), 69–93.
- Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1), 61–74.
- Edwards, J. A. & Lampert, M. D. (1993). *Talking Data: Transcription and Coding in Discourse Research*. Lawrence Erlbaum Associates.
- Evert, S., & Hardie, A. (2011, July). Twenty-first century corpus workbench: updating a query architecture for the new

- millennium. Proceedings of the Corpus Linguistics 2011 conference içinde (ss. 1-21).
- Francis, W. N., & Kučera, H. (1964). *Manual of information to accompany a standard corpus of present-day edited American English, for use with digital computers*. Providence, RI: Brown University.
- Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Houghton Mifflin.
- Gaspari, F. (2022). Expanding the reach of corpus based translation studies, the opportunities that lie ahead. Granger, S. & Lefer, M.-A. (Ed.), *Extending the Scope of Corpus-Based Translation Studies* içinde (ss. 42-60). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Granger, S., & Lefer, M.-A. (Ed.). (2022). *Extending the Scope of Corpus-Based Translation Studies*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Greenbaum, S. (1996). *Comparing English worldwide: The International Corpus of English*. Oxford University Press.
- Gries, S. T. (2009). *Statistics for linguistics with R: A practical introduction*. Berlin: Mouton de Gruyter.
- Groom, N. (2013). COBUILD Project. C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics*. içinde, Wiley-Blackwell.
- Hoffmann, S.; Evert, S., Smith, N. ve Berglund Prytz, Y. (2008). *Corpus Linguistics with BNCweb – A Practical Guide*. Peter Lang.
- Holmes, J. S. (1988). *Translated! Papers on literary translation and translation studies*. Rodopi.
- Hu, K., & Kim, K. H. (2024). Corpus-Based Translation Studies. The Routledge Handbook of the History of Translation Studies içinde (ss. 293-308). Routledge.
- Huddleston, R., & Pullum, G. K. (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

- IBM Corp. (2020). *IBM SPSS Statistics for Windows* (Sürüm 27.0) [Bilgisayar yazılımı]. Armonk, NY: IBM Corp.
- Ide, N. & Romary, L. (2004). A Registry of Standard Data Categories for Linguistic Annotation. Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC) içinde.
- Ide, N., & Suderman, K. (2004). The American National Corpus first release. In Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC) içinde (ss. 1681–1684).
- Järvinen, T. (t.y.). The Bank of English: Morphosyntactic annotation. Helsinki Üniversitesi, Dilbilim ve Bilgisayar Destekli Dilbilim Birimi. <http://www2.lingsoft.fi/doc/engcg/Bank-of-English.html>
- Johansson, S., Leech, G., & Goodluck, H. (1978). *Manual of information to accompany the Lancaster–Oslo/Bergen Corpus of British English, for use with digital computers*. Department of English, University of Oslo.
- Johansson, S. (2002). Parallel corpora in linguistics. S. Johansson & S. Oksefjell (Ed.), *Corpora and cross-linguistic research* içinde (ss. 3–24). Rodopi.
- Kennedy, G. (1998). *An introduction to corpus linguistics*. Longman.
- Kennedy, G. (2014). *An introduction to corpus linguistics*. London: Routledge.
- Kilgarriff, A., Rychlý, P., Smrž, P., & Tugwell, D. (2004). The Sketch Engine. *Proceedings of the 11th EURALEX International Congress* içinde (ss. 105–116).
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1(1), 7–36.
- Knowles, Gerry (1996). *Spoken English on Computer: Transcription, Mark-up and Application*. London: Longman.

- Kruger, K. Wallmach & J. Munday (Ed.), *Corpus-based translation studies: Research and applications* içinde (ss. 33–75). London: Continuum.
- Küçük, M., & Can, F. (2008). TamBiC: A bilingual concordancer for translation studies. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)* içinde (ss. 2335–2339). Marrakech: ELRA.
- Laviosa, S. (1998). The corpus-based approach: A new paradigm in translation studies. *Meta: Translators' Journal*, 43(4), 474–479.
- Laviosa, S. (2002). *Corpus-based Translation Studies: Theory, findings, applications*. New York: Rodopi.
- Leech, G. (1992). Corpus linguistics and theories of linguistic performance. J. Svartvik (Ed.), *Directions in corpus linguistics: Proceedings of Nobel Symposium 82* içinde (ss. 105–122). Mouton de Gruyter.
- Leech, G. (1997). Introducing corpus annotation. R. Garside, G. Leech & T. McEnery (Ed.), *Corpus annotation: Linguistic information from computer text corpora* içinde (ss. 1–18). London: Longman
- Leech, G., Myers, G. & Thomas, J. (1995). *Spoken English on Computer: Transcription, Mark-up and Application*. London: Longman.
- Mahadi, T. S. T., Vahedi, V., & Norizan, A. R. (2010). A framework for developing a do-it-yourself corpus. *Journal of Language Studies*, 10(2), 1–16.
- Malmkjær, K. (1998). Parallel corpora. M. Baker (Ed.), *Routledge encyclopedia of translation studies* içinde (ss. 534–537). London: Routledge.
- Matsubara, S., Horiguchi, S., Kikui, G., Sakai, S., Paul, M., & Ney, H. (2002). Automatic speech recognition for simultaneous interpretation. *Proceedings of the International Conference on Spoken Language Processing (ICSLP)* içinde (ss. 2681–2684).
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.

- Metzger, M., & Roy, C. (2011). Discourse and interpreting. K. Malmkjær & K. Windle (Ed.), *The Oxford handbook of translation studies* içinde (ss. 423–437). Oxford: Oxford University Press.
- Meyer, C. F. (2002). *English corpus linguistics: An introduction*. Cambridge, UK: Cambridge University Press.
- Mikhailov, M., & Cooper, R. (2016). *Corpus linguistics for translation and contrastive studies: A guide for research*. London: Routledge.
- Monti, C., Bendazzoli, C., Sandrelli, A., & Russo, M. (2005). The European Parliament Interpreting Corpus (EPIC): Methodology and preliminary results. *Interpreting*, 7(2), 155–170.
- O'Brien, S. (2012). Translation as human–computer interaction. *Translation Spaces*, 1(1), 101–122.
- Ono, T., Horiguchi, S., Matsubara, S., & Kikui, G. (2008). A study of disfluencies in simultaneous interpreting. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC)* içinde (ss. 125–132).
- Pastor, G. C., & Seghiri, M. (2009). Virtual corpora as resources for translators. *Meta: Translators' Journal*, 54(4), 729–748.
- Plungian, V., & Biryukov, A. (2008). Russian National Corpus: Structure and development. N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis & D. Tapias (Eds.), *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. European Language Resources Association (ELRA). içinde
- R Core Team. (2023). *R: A language and environment for statistical computing* [Bilgisayar yazılımı]. Vienna: R Foundation for Statistical Computing.
- Rouse, R. H., & Rouse, M. A. (1979). *Preachers, florilegia and sermons: Studies on the manipulation of texts*. Toronto: Pontifical Institute of Mediaeval Studies.
- Sager, J. C. (1990). *A practical course in terminology processing*. Amsterdam: John Benjamins.

- Salkie, R. (2002). How can linguists profit from parallel corpora?. *Language and Computers*, 43, 93-110.
- Scardulla, S. (2016). Interpreting at the European Commission: A corpus-based study. *The Interpreters' Newsletter*, 21, 7-26.
- Setton, R. (2011). Corpus-based interpreting studies (CIS): Overview and prospects. A. Kruger, K. Wallmach, & J. Munday (Ed.), *Corpus-based translation studies: Research and applications içinde* (ss. 33-75). Continuum International.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Scott, M. (2008). WordSmith tools version 5. Liverpool: Lexical Analysis Software.[Bilgisayar Yazılımı]
- Somers, H. (2003). Translation memory systems. *Computers and Translation: A Translator's Guide*, 31-48.
- Sperberg-McQueen, C. M., & Burnard, L. (Ed.). (2002). *Guidelines for electronic text encoding and interchange (TEI P4)*. Chicago: Text Encoding Initiative Consortium.
- Smith, D., & Baker, J. (2006). The Lampeter Corpus of Early Modern English Tracts. Oxford: Oxford Text Archive.
- Spevack, M. (1968). *A complete and systematic concordance to the works of Shakespeare*. Hildesheim: Georg Olms.
- Summers, D. (Ed.). (1987). *Longman dictionary of contemporary English* (2nd ed.). London, UK: Longman.
- Svartvik, J. (Ed.). (1990). \*The London-Lund Corpus of Spoken English: Description and Research\*. Lund: Lund University Press.
- Svartvik, J. (1992). Corpus linguistics comes of age. *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82 Stockholm, 4-8 August 1991 içinde* (ss. 7-16).
- Svartvik, Jan & Quirk, Randolph (1980). *A Corpus of English Conversation*. Lund Studies in English, Gleerup.
- TEI Consortium. (2023). *TEI P5: Guidelines for Electronic Text Encoding and Interchange (Version 4.7.0) [Guidelines]*. Zenodo.

- Teubert, W. (1996). Comparable or parallel corpora? *International Journal of Lexicography*, 9(3), 238–264.
- Thomas, M., & Short, M. (1994). *Using concordances on personal computers: A guide to the principles and practice*. London: Longman.
- Tiedemann, J. (2016). OPUS – Parallel corpora for everyone. *Baltic Journal of Modern Computing*, 4(2), 384–403.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins.
- Tohyama, H., Matsubara, S., Horiguchi, S., & Kikui, G. (2005). Constructing a simultaneous interpreting corpus for speech translation research. *Proceedings of Interspeech 2005* içinde (ss. 277–280).
- Toury, G. (1995). *Descriptive translation studies and beyond*. John Benjamins.
- Toutanova, K., Klein, D., Manning, C. D., & Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)* içinde (ss. 173–180).
- Unicode Consortium. (2023). *The Unicode Standard, Version 15.1.0*. South San Francisco, CA: Unicode Consortium.
- Varantola, K. (2003). Translators and disposable corpora. F. Zanettin, S. Bernardini & D. Stewart (Ed.), *Corpora in translator education* içinde (ss. 55–70). Manchester: St. Jerome Publishing.
- Variation, Contacts and Change in English Araştırma Grubu. (t.y.). *VARIENG: Variation, Contacts and Change in English*. Helsinki Üniversitesi. Erişim tarihi: 24 Ekim 2025, <https://varieng.helsinki.fi/>
- Wang, B. (2015). Corpus-based interpreting studies: The Chinese perspective. *Perspectives: Studies in Translatology*, 23(1), 111–127.
- Zanettin, F. (2012). *Translation-driven corpora: Corpus resources for descriptive and applied translation studies*. Manchester: St. Jerome Publishing.

# Ek-A

## Terimce (Glossary)

### A

#### **Ampirik Araştırma (Empirical Research)**

Gözleme, ölçüme ve gerçek dil verilerine dayanan araştırma yaklaşımı. Derlem temelli çalışmalar ampirik araştırmaların temel örneklerindedir.

#### **Anotasyon (Annotation)**

Derlemdeki metinlere dilbilgisel, sözcüksel, anlamsal veya söylemsel bilgilerin etiketleme yoluyla eklenmesi işlemi.

---

### B

#### **Betimleyici Çeviri Çalışmaları (Descriptive Translation Studies – DTS)**

Çeviriyi “nasıl olması gerektiği” açısından değil, “gerçekte nasıl gerçekleştiği” açısından inceleyen, hedef odaklı ve ampirik çeviri araştırma yaklaşımı.

#### **İki Dilli Derlem (Bilingual Corpus)**

İki farklı dilde metinler içeren derlem türü; paralel veya karşılaştırılabilir olabilir.

---

### C

#### **Derlem (Corpus)**

Belirli ilkeler doğrultusunda seçilmiş, bilgisayarca okunabilir, sözlü ve/veya yazılı metinlerden oluşan yapılandırılmış metin koleksiyonu.

#### **Derlem Tabanlı Yaklaşım (Corpus-based Approach)**

Dilsel analizlerin sezgiye değil, büyük ölçekli gerçek dil verilerine dayandırıldığı araştırma yaklaşımı.

## **Translation Corpus (Çeviri Derlemi)**

Aynı dilde yazılmış orijinal ve çeviri metinleri içeren tek dilli derlem türü.

---

D

### **Derlem Büyüklüğü (Corpus Size)**

Bir derlemde yer alan toplam kelime sayısı. Araştırma amacına göre küçük veya büyük ölçekli olabilir.

### **Derlem Tasarımı (Corpus Design)**

Derlemin amaç, temsil edilebilirlik, metin türü, dil, dönem ve büyüklük gibi kriterlere göre planlanması süreci.

---

E

### **Eşdizimlilik (Collocation)**

Belirli sözcüklerin istatistiksel olarak birlikte ortaya çıkma eğilimi.

### **Açıklık (Explicitness)**

Çeviri metinlerde bilgilerin daha açık ve belirgin hâle getirilmesi eğilimi.

---

H

### **Hizalama (Alignment)**

Paralel derlemlerde kaynak ve hedef metinlerin cümle, tümce veya paragraf düzeyinde eşleştirilmesi işlemi.

---

K

### **Karşılaştırılabilir Derlem (Comparable Corpus)**

Aynı dilde veya farklı dillerde, benzer metin türlerinden oluşan ancak çeviri ilişkisi bulunmayan metin derlemi.

### **Keyword Analysis (Anahtar Sözcük Analizi)**

Bir derlemde belirli sözcüklerin referans derleme kıyasla istatistiksel olarak anlamlı biçimde öne çıkmasını inceleyen yöntem.

### **Bağlamda Anahtar Sözcük Gösterimi, KWIC (Key Word in Context)**

Anahtar sözcüğün sağ ve sol bağlamıyla birlikte gösterildiği bağlam dizini formatı.

---

M

### **Tek Dilli Derlem (Monolingual Corpus)**

Tek bir dilde metinler içeren derlem türü.

### **Çok Dilli Derlem (Multilingual Derlem)**

İkiden fazla dilde metinler içeren derlem türü; paralel veya karşılaştırılabilir olabilir.

---

N

### **Normalleştirme (Normalization)**

Çeviri metinlerin hedef dilin yerleşik normlarına daha fazla uyum göstermesi eğilimi.

---

O

### **OCR (Optical Character Recognition – Optik Karakter Tanıma)**

Basılı metinlerin taranarak dijital ve düzenlenebilir metin hâline getirilmesini sağlayan teknoloji.

---

P

### **Paralel Derlem (Parallel Corpus)**

Kaynak metinler ve bunların çevirilerinden oluşan, hizalanmış çok dilli derlem türü.

### **Kuralcı Yaklaşım (Prescriptive Approach)**

Çevirinin nasıl yapılması gerektiğini belirlemeye çalışan normatif yaklaşım.

---

## **S**

### **Sadelik (Simplification)**

Çeviri metinlerde dilin daha basit ve daha az karmaşık hâle gelmesi eğilimi.

### **Sözcük Sıklığı (Word Frequency)**

Bir sözcüğün derlem içinde kaç kez geçtiğini gösteren ölçüt.

---

## **T**

### **Temsil Edilebilirlik (Representativeness)**

Bir derlemin, temsil ettiği dili veya metin türünü ne ölçüde yansıttığını ifade eden ilke.

### **Translation Universals (Çeviri Evrenselleri)**

Dil çiftinden bağımsız olarak çeviri metinlerde ortaya çıktığı varsayılan ortak dilsel eğilimler.

---

## **V**

### **Veriye Dayalı Çeviribilim (Data-driven Translation Studies)**

Çeviri olgusunu derlem verileri üzerinden inceleyen ampirik araştırma yaklaşımı.

# EK - B

## ALIŐTIRMALAR

---

### *Bölüm I: Temel Kavramlar*

---

#### 1. Derlem Dilbilimine Giriş

Farklı derlem türlerini tanımlayın ve her biri için gerçek dünyadan örnekler bulun.

İnternette bir açık erişimli derlem (örneğin BNC, COCA) seçin ve ana sayfasındaki kullanım özelliklerini özetleyin.

#### 2. Çeviribilime Giriş

Aşağıdaki kısa metni iki farklı çeviri stratejisi kullanarak çevirin (örn. birebir çeviri – anlam odaklı çeviri). Çevirinizde karşılaştığınız kültürel ve dilsel zorlukları analiz edin.

#### 3. Derlem ve Çeviribilim Arasındaki İlişki

Bir çeviri örneği bulun (örneğin bir roman çevirisi) ve bunu bir paralel derlem yardımıyla analiz etmeyi planlayın: Hangi yönleri incelemek istersiniz? Hangi araçları kullanırdınız?

#### 4. Derlem Oluşturma

Kendi mini derlemınızı oluşturun (en az 10.000 kelime): bir alana odaklanın (ör. yemek tarifleri, film eleştirileri, akademik makaleler).

Derlemınızı metin türüne göre etiketleyin ve üstveri dosyası hazırlayın.

#### 5. Derlem Türlerini Karşılaştırma

Aynı konuda yazılmış Türkçe ve İngilizce haber metinlerinden birer karşılaştırmalı derlem oluşturun.

Hangi sözcük kalıpları benzer, hangileri farklı? Kısa bir analiz yapın.

## 6. Yazılım Kullanımı

AntConc kullanarak küçük bir metin derlemi üzerinde:

En sık geçen 20 sözcüğü listeleyin.

“environment” sözcüğü için bağlam dizini listesi çıkarın.

“take” fiilinin birlikte en sık geçtiği sözcükleri analiz edin.

## 7. Çevirmen Eğitiminde Derlem Kullanımı

Alıştırma:Daha önce çevrilmiş kısa bir metni alın. AntConc ile anahtar kelimeleri çıkararak metindeki ana temaları belirleyin. Aynı metni tekrar çevirin ve farkları karşılaştırın.

## 8. Çeviri Evrenselleri

Bir orijinal metin ve çevirisini karşılaştırarak:

Nerelerde açıklayıcılık açıklama yapılmış?

Nerelerde sadeleştirme var?

Tablolarla gösterin ve yorumlayın.

## 9. Kalite Değerlendirmesi

İki farklı çeviri versiyonunu karşılaştırarak kalite farklarını analiz edin.

Hangi versiyon normlara daha uygun? Gerekçeleriyle açıklayın.

## 10. Karşıtsal Analiz

Türkçe'deki deyimlerin İngilizce karşılıklarını araştırın (örn. “ateş bacayı sardı”).

Seçtiğiniz 5 deyim derlem kullanarak çevirilerde nasıl yer aldığını gösterin.

## 11. Uzmanlık Alanlarında Çeviri

Tıbbi bir metin seçin ve terim listesi çıkarın.

Bu terimleri derlem aracılığıyla hedef dilde nasıl çevrildiğini inceleyin.

## 12. Görsel-işitsel Çeviri

Netflix/YouTube altyazısı ile orijinal diyalogları karşılaştırın. Bir dizi sahnesinde çevirmen tercihlerini derlem tabanlı bir yaklaşımla analiz edin.

## 13. Makine Çevirisi

Bir metni Google Translate ve DeepL ile çevirin. Her iki çeviriyi karşılaştırın ve bir derlem yardımıyla hataları sınıflandırın.

## 14. Sözlü Çeviri Çalışmaları

Bir konferans çevirisi videosunu dinleyin. Duraksamalar, dolaylı ifadeler ve özetlemeler üzerine notlar alın. Bunları sözlü çeviri derlemlerinden örneklerle karşılaştırın.

## 15. Güncel Eğilimler

Derlem-odaklı bir çeviri araştırması makalesi bulun.

## Ek C

Biber etiketleyici (Biber Tagger) tarafından atanan etiketlerin listesi

### Açıklama

Beş etiket alanı vardır ve bunlar artı işaretiyle (+) ayrılmıştır. Çoğu kelime için sadece bir veya iki alan kullanılır. Bir kelimenin birincil dilbilgisi kategorisi genellikle ilk etiket alanında işaretlenir; bu ilk alan etiketlerinin birçoğu CLAWS etiket setinde kullanılan etiketlerle aynıdır (Garside, Leech, & Sampson, 1987).

Sıfatlar (TAG = JJ), isimler (TAG = NN) ve fiiller (TAG = VB) durumunda, etiket

?? Etiket Alanı 4'te sözlükte olmayan kelimeleri işaretlemek için görünebilir; Bu durumlarda, dilbilgisi kategorisi morfoloji ve çevresindeki bağlama göre atanır.

Etiket Alanı 4'teki xvbn ve xvbnx etiketleri, işlev fark etmeksizin bir kelimeyi geçmiş zaman ortacı (past participle) olarak işaretler. Bu nedenle, bazı sıfatlar, isimler ve temel fiil formları xvbn olarak işaretlenir. Tüm geçmiş zaman fiilleri, bitmişlik görünüşlü fiiller ve pasif fiiller bu etiketi taşır. xvbnx etiketi, dilbilgisel fonksiyonun (örneğin, bitmişlik görünüşü veya pasif) bağlamdan çok yüksek doğrulukla tanımlandığı durumlar için kullanılır; Xvbn etiketi, atanan dilbilgisel fonksiyonun daha az kesin olduğu durumlar için kullanılır.

Etiket Alanı 4'teki xvbg ve xvbgx etiketleri, fonksiyon fark etmeksizin bir kelimeyi şimdiki zamanki ortağ biçimi olarak işaretler. Bu nedenle, bazı sıfatlar ve isimler xvbg olarak işaretlenir. Mevcut tüm süreklilik gösteren fiiller bu etikete sahiptir. xvbgx etiketi, dilbilgisi fonksiyonunun çok yüksek doğrulukla tanımlandığı durumlar için kullanılır; Xvbg

etiketi, atanan dilbilgisel fonksiyonun daha az kesin olduğu durumlar için kullanılır.

---

Etiket Dizisi: Alan 1 + Alan 2 + Alan 3 + Alan 4 + Alan 5

:+clp+++ colon + clause punctuation

;/+clp+++ semi-colon + clause punctuation

?+clp+++ question mark + clause punctuation

!+clp+++ exclamation mark + clause punctuation

,++++ comma

-++++ dash

"++++ double quote mark '+++++ single quote mark (++++ left parenthesis

)++++ right parenthesis

\$++++ dollar sign

%++++ percent sign &fo++++ formula symbols

&fw++++ foreign word

abl++++ pre-qualifier (rather, such) abn++++ pre-quantifier (all, half)

abx++++ pre-quantifier/double conjunction (both)

ap++++ post-determiner (many, more, most, only, other, own, same, ...) aps++++ (others)

at++++ singular indefinite article (a, an) ati++++ singular definite article (the, no)

cc++++ coordinating conjunction (and, but, or)

cc+cls+++	coordinating conjunction + clausal connector
cc+phrs+++	coordinating conjunction + phrasal connector
cc''+++	multi-word coordinating conjunction (as well as)
cc++neg++	coordinating conjunction + + negation (nor)
cd+++	cardinal number (2, 3, 4, two, three, four, hundred, ...)
cd+date+++	cardinal number + date (year only)
cdl+++	cardinal number: 1, one
cdls+++	cardinal number: ones
cds+++	cardinal plural (tens, hundreds, thousands)
od+++	ordinal number (1st, 2nd, first, second, ...)
cs+cnd+++	subordinating conjunction + conditional (if, unless)
cs+con+++	subordinating conjunction + concessive (although, though)
cs+cos+++	subordinating conjunction + causative (because)
cs+who+++	subordinating conjunction + WH word (whether)
cs+sub+++	subordinating conjunction + other (as, except, until, ...)
cs''+++	multi-word subordinating conjunction (in that, so that, ...)
dt+dem+++	determiner + demonstrative (this, that, these, those modifying N)
dt+pdem+++	determiner + demonstrative pronoun (this, that, these, those)
dti+++	singular or plural determiner (any, enough, some)
dt+++	other singular determiner (another, each)
dtx+++	determiner/double conjunction (either)
ex+pex+++	existential there
in+++	preposition
in+ppvb+++	preposition + prepositional verb (account for,

join in, ...) in+pl+++	preposition + place marker (above, behind, beside, ...) in"+++	multi-word preposition (as to, away from, instead of, ...) in+strn+++	preposition + stranded
jj+atr+++	adjective + attributive function		
jj+atr+++xvbg+	adjective + attributive function + + -ing form		
jj+atr+++xvbn+	adjective + attributive function + + past participle form		
jj+pred+++	adjective + predicative function		
jj+++	adjective + indeterminate function		
jjb+atr+++	attributive-only adjective + attributive		
(chief, entire) jjr+atr+++	comparative adjective + attributive function		
jjr+pred+++	comparative adjective + predicative function		
jjt+atr+++	superlative adjective + attributive function		

Tüm modal formlar, Alan 5'te 0 olarak işaretlenebilir (örneğin, md+prd+++ 0) ve böylece kısaltılmış biçimler olduklarını gösterebilir (örneğin, 'll, 've)

md+nec+++	modal + necessity (ought, should, must)
md+pos+++	modal + possibility (can, may, might, could)
md+prd+++	modal + prediction (will, would, shall)
md"++pmd"++	modal + + multi-word periphrastic modal (e.g., be going to)

nn+++	singular common noun
nn+nom+++	singular noun + nominalization
nvbg+++xvbg+	singular noun + + + -ing form
nn+++xvbn+	singular noun + + + past participle form
nns+++	plural common noun

nns+nom+++	plural noun + nominalization nnu++++ unit of measurement (lb, kg, ...) np++++ singular proper noun
nps+++	plural proper noun
npl+++	locative noun
npt+++	singular titular noun
npts+++	plural titular noun
nr++++	singular adverbial noun (east, west, to- day, home, ...) nrs++++ plural adverbial noun

Not: Aşağıdaki zamir etiketlerinde, birinci şahıs karakterini belirtmek için 1 harfi kullanılırken, l harfi (yani küçük L) dö-  
nümlü yapıları işaretlemek için kullanılır.

pp1a+pp1+++	first person subject pronoun + first per- son pronoun pp1a+pp1+++0 first person subject pronoun + 1st person pro. + contracted pp1o+pp1+++ first person ob- ject pronoun + first person pronoun pp\$+pp1+++ posses- sive determiner + first person pronoun (my, our) ppl+pp1+++ singular reflexive pronoun + first person pronoun (myself) ppls+pp1+++plural reflexive pronoun + first person pronoun (ourselves) pp2+pp2+++ second person pronoun + second person pronoun
pp\$+pp2+++	possessive determiner + second person pronoun (your) ppl+pp2+++ singular reflexive pronoun + second person pronoun (yourself) pp3a+pp3+++ third person subject pronoun + third person personal pronoun pp3o+pp3+++ third person object pronoun + third per- son personal pronoun pp3+pp3+++0 third person pronoun + 3rd person personal pro. + contracted pp\$+pp3+++

possessive + 3rd pers. personal pro. (his, her, their) ppl+pp3+++ sg. reflexive pronoun + 3rd pers. personal pro. (her/himself) ppls+pp3+++ pl. reflexive pronoun + 3rd pers. personal pro. (themselves) pp3+it+++ third person pronoun + third person impersonal pronoun (it) pp\$+it+++ possessive determiner + third person impersonal pronoun (its) pp\$\$++++ possessive pronoun (mine, yours, ...)

pn"++++ multi-word nominal pronoun (no one, ...)  
 pn++++ nominal pronoun (someone, everything, ...)

ql++++ qualifier + (as, less, more, too) ql+amp+++ qualifier + amplifier (very) ql+emph+++ qualifier + emphatic (most) qlp++++ post-qualifier (enough, indeed)

Tüm zarf formları, Alan 3'te 'spl't' olarak işaretlenebilir (örneğin, rb+amp+spl't) ve zarfın yardımcı karakter içinde geçtiğini belirtmek için

(Örneğin, they've probably been looking...).

rb++++ general adverb

rb"++++ multi-word adverb (at last, in general)

rb+cnj adverb + conjunct (however, therefore, thus, ...) rb++neg++ neither

rb+amp+++ adverb + amplifier (absolutely, completely, entirely, ...) rb+down+++ adverb + downtoner (nearly, only, merely, ...)

rb+emph+++ adverb + emphatic (just, really, so, ...)

rb+hdg+++ adverb + hedge (almost, maybe, ...)

rb"+hdg"+++ multi-word adverb + hedge (kind of, sort

of) rb+phrv+++ adverb + phrasal verb (get in, wrap up, ...)

rb+pl+++	adverb + place marker (abroad, ahead, far, upstream, ...)
rb+tm+++	adverb + time marker (afterwards, again, immediately, ...)
rb+dspt+++	adverb + discourse particle (anyway, well, ...)
rbr++++	comparative adverb (better, quicker)
rbr+tm+++	comparative adverb + time marker (earlier, later, sooner, ...)
rn+pl+++	nominal adverb + place marker (here, there)
rn+tm+++	nominal adverb + time marker (now, then)
rn+dspt+++	nominal adverb + discourse particle (now)
rp++++	adverbial particle (back, in, round, up, ...)
rp+pl+++	adverbial particle + place marker (away, behind, out, ...)

---

tht+jcmp+++	that as dependent clause head + adjective complement
tht+ncmp+++	that as dependent clause head + noun complement
tht+vcmp+++	that as dependent clause head + verb complement
tht+rel+++	that as dependent clause head + relative clause
tht+rel+obj++	that as dep. clause head + relative clause + object position
tht+rel+subj++	that as dep. clause head + relative clause + subject position

---

to++++	infinitive marker
to"++++	multi-word infinitive marker (in order to)
uh++++	interjection/filler (hey, oh, ok, yes, erm ...)
vb++++	base form of verb, excluding verbs in infinitive clauses (uninflected present tense, imperative)
vb+++xvbn+	base form of verb + + + past participle

form

(e.g., cut, hit, hurt, ...) vb+be+aux++ base form of verb + be  
+ auxiliary verb vb+be+vrb++ base form of verb + be + main

verb vb+bem+aux++ verb + am + auxiliary verb

vb+bem+aux++0 verb + am + auxiliary verb ++ contracted

(m) vb+bem+vrb++ verb + am + main verb

vb+bem+vrb++0 verb + am + main verb ++ contracted (m)

vb+ber+aux++ verb + are + auxiliary verb

vb+ber+aux++0 verb + are + auxiliary verb ++ contracted

(re) vb+ber+vrb++ verb + are + main verb

vb+ber+vrb++0 verb + are + main verb ++ contracted (re)

vb+do+aux++ verb + do + auxiliary verb

vb+do+vrb++ verb + do + main verb vb+hv+aux++

verb + have + auxiliary verb

vb+hv+aux++0 verb + have + auxiliary verb ++ con-  
tracted (ve) vb+hv+vrb++ verb + have + main verb

vb+hv+vrb++0 verb + have + main verb ++ contracted

(ve) vb+seem+++ base form of verb + seem/appear

vb+vprv+++ base form of verb + private verb (believe,  
feel, think, ...) vb+vprv+tht0++ base form of verb + private

verb + that deletion \*\* vb+vpub+++ base form of verb +

public verb (assert, complain, say, ...) vb+vpub+tht0++ base

form of verb + public verb + that deletion \*\* vb+vsua+++

base form of verb + suasive verb (ask, command, insist, ...)

---

Tüm geçmiş zaman fiil formları (were, was, did, had hariç)  
ya xvbn ya da xvbnx olarak işaretlenir; Aşağıda yalnızca  
xvbn içeren diziler listelenmiştir.

vbd+++xvbn+	past tense verb + + + past participle form
vbd+bed+aux++	past tense verb + were + auxiliary verb
vbd+bed+vrb++	past tense verb + were + main verb
vbd+bedz+aux++	past tense verb + was + auxiliary verb
vbd+bedz+vrb++	past tense verb + was + main verb
vbd+dod+aux++	past tense verb + did + auxiliary verb
vbd+dod+vrb++	past tense verb + did + main verb
vbd+hvd+aux++	past tense verb + had + auxiliary verb
vbd+hvd+vrb++	past tense verb + had + main verb
vbd+seem++xvbn+	past tense verb + seem/appear
vbd+vprv++xvbn+	past tense + private verb
(believe, feel, think, ...) vbd+vprv+tht0+xvbn+	past tense + private verb + that deletion **
vbd+vpub++xvbn+	past tense + public verb (assert, complain, say, ...)
vbd+vpub+tht0+xvbn+	past tense + public verb + that deletion **
vbd+vsua++xvbn+	past tense + suasive verb (ask, command, insist, ...)

\*\* THT0 etiketi, tamamlayıcı sildiğinde takip eden bir tamamlayıcı cümlesinin oluşmasını işaret eder

Mevcut tüm süreklilik bildiren fiil formları ya xvbg ya da xvbgx olarak işaretlenir; Aşağıda yalnızca xvbg içeren diziler listelenmiştir.

vbg+++xvbg+	present progressive verb + + + -ing form
vbg+beg++xvbg+	present progressive verb + being
vbg+beg+aux+xvbg+	present progressive verb + being + auxiliary verb
vbg+hvg++xvbg+	present progressive verb + having
vbg+vprv++xvbg+	pres. prog. + private verb (believe, feel, think, ...)
vbg+vprv+thtO+xvbg+	present progressive + private verb + that deletion **
vbg+vpub++xvbg+	pres. prog. +

public verb (assert, complain, say, ...) vbg+vpub+thtO+xvbg+  
present progressive + public verb + that

deletion\*\* vbg+vsua++xvbg+ pres. prog. + suasive verb  
(ask, command, insist, ...)

vwbg+++xvbg+ present progressive postnominal modi-

fier vwbg+beg++xvbg+ present progressive post-

nominal modifier + being vwbg+hvg++xvbg+ present pro-

gressive postnominal modifier + having vwbg+vprv++xvbg+

present prog. postnom. modifier + pri-

vate verb vwbg+vpub++xvbg+ present prog. postnom.

modifier + public verb

vbi++++ base form of verb in infinitive clause

vbi+vprv+++ infinitive verb + private verb (believe,

feel, think, ...) vbi+vprv+tht0++ infinitive verb + private verb

+ that deletion \*\* vbi+vpub+++ infinitive verb + public verb

(assert, complain, say, ...) vbi+vpub+tht0++ infinitive verb +

public verb + that deletion \*\* vbi+vsua+++ infinitive verb

+ suasive verb (ask, command, insist, ...)

vbz +++ 3rd person singular verb

vbz+bez+aux++ 3rd person sg. verb + is + auxiliary verb

vbz+bez+aux++0 3rd person sg. + is + auxiliary verb + +

contracted (s) vbz+bez+vrb++ 3rd person sg. verb + is + main  
verb

vbz+bez+vrb++0 3rd person sg. + is + main verb + +

contracted (s) vbz+doz+aux++ 3rd person sg. verb + does +

auxiliary verb vbz+doz+vrb++ 3rd person sg. verb + does +

main verb

vbz+hvz+aux++	3rd per- sg.verb + has + auxiliary verb son
vbz+hvz+vrb++	3rd per- sg.verb + has + main verb son
vbz+seem+++	3rd per- sg.verb + seem/appear son
vbz+vprv+++	3rd per- sg.+ private verb (believe, feel, ...) son think,
vbz+vprv+tht0++	3rd per- sg.+ private verb + that deletion son **
vbz+vpub+++	3rd per- sg.+ public verb (assert, com- ...) son plain, say,
vbz+vpub+tht0++	3rd per- sg.+ public verb + that deletion son **
vbz+vsua+++	3rd per- sg.+ suasive verb (ask, com- ...) son mand, insist,

\*\* THT0 etiketi, tamamlayıcı sildiğinde takip eden bir tamamlayıcı cümlesinin oluşmasını işaret eder

All perfect aspect verb forms and passive verb forms are marked as either xvbn or xvbnx; only the sequences with xvbn are listed below.

vprf+++xvbn+	perfect aspect verb + + + past participle form
vprf++tht0+xvbn+	perfect aspect verb + + that deletion **
vprf+ben+aux+xvbn+	perfect aspect verb + been + auxiliary verb
vprf+ben+vrb+xvbn+	perfect aspect verb + been + main verb
vpsv++agls+xvbn+	main clause passive verb + + agentless
vpsv++by+xvbn+	main clause passive verb + + by passive

vwb++++xvbn+ passive postnominal modifier +++ past participle form vwb+vprv++xvbn+ passive postnominal modifier + private verb vwb+vpub++xvbn+ passive postnominal modifier + public verb vwb+vsua++xvbn+ passive postnominal modifier + suasive verb

---

wdt+who+++ WH determiner+WH word (what, whatever, whichever, ...) wdt+who+whcl++ WH determiner + WH word + WH clause

wdt+who+whq++ WH determiner + WH word + WH question

whp+rel+obj++ WH pronoun + relative clause + object position whp+rel+pied++ WH pronoun + relative clause + object position with

prepositional fronting (pied piping) whp+rel+subj++ WH pronoun + relative clause + subject position

whp+who+++ WH pronoun + WH word (not a relative clause) whp+who+whq++ WH pronoun + WH word + WH question

wrb+who+++ WH adverb (how, when, where, ...) + WH word wrb+who+whcl++ WH adverb + WH word + WH clause wrb+who+whq++ WH adverb + WH word + WH question

xnot++not++ not ++ negation

xnot++not++0 not ++ negation ++ contracted form (nt)

xvbn+++xvbn+ past participle form - indeterminate grammatical function xvbg+++xvbg+ present participle form - indeterminate grammatical function

zz++++ letter of the alphabet