

Araştırma Makalesi / Research Paper

HALAÇ TÜRKÇESİNİN AĞIZLARI İLE GÜNEY AZERBAYCAN TÜRKÇESİNİN KUM (COUZEH) AĞZI ARASINDAKİ SÖZLÜKSEL UZAKLIK: LEVENSHTAIN VE DIJKSTRA ALGORİTMALARI AMALGAMI

Mehmet AKKUŞ*

İbrahim GÖKCAN**

Öz

Bu çalışma, İran'ın Merkezi ve Kum vilayetlerinde konuşulmakta olan iki Türk dili değişkesinin Halaç Türkçesi ve Güney Azerbaycan Türkçesi- sözlüksel uzaklığını Leipzig-Jakarta çekirdek sözcük listesindeki maddeler temelinde saptamayı amaçlamaktadır. Bu bağlamda Halaç Türkçesinin Merkezi ve Kum vilayetlerinde konuşulan, Doerfer'in tasnif ettiği olduğu tüm ağız bölgeleri ile Güney Azerbaycan Türkçesinin Kum'da konuşulduğu tanıklanmış olan Couzeh değişkesi, çalışmada karşılaştırmalı olarak incelenecektir. Yöntem olarak nicel lehçe/ağız ölçüm yöntemleri arasında ilgili uluslararası alan yazında sıklıkla yararlanılan Levenshtein uzaklık algoritması (LUA) söz konusu değişkeler arasındaki uzaklığı saptamak amacıyla kullanılacak, sonrasında ortaya çıkan değerler ışığında Güney Azerbaycan Türkçesinin (GAT) Couzeh değişkesi ile Halaç türkçesi ağızları arasında Dijkstra algoritması ile minimum yol sorgulanacaktır. Dijkstra algoritmasının uygulanmasında ağızlar birer köşe ve ağızlar arası LUA ile ölçülen değerler köşeler arası kenar değerleri olarak alınacaktır. Çalışmada ayrıca eskicil özellikleri ile tanınan ve Türk dilinin *sui generis* bir dalını teşkil ettiği belirtilen Halaç Türkçesi (Tekin, 1989) ile bir Oğuz

Geliş Tarihi/Date Applied: 24.02.2023

Kabul Tarihi/Date Accepted: 26.05.2023

Makalenin Künyesi: Akkuş, M. ve Gökcan, İ. (2023). Halaç Türkçesinin ağızları ile Güney Azerbaycan Türkçesinin Kum (Couzeh) ağızı arasındaki sözlüksel uzaklık: Levenshtein ve Dijkstra algoritmaları amalgamı. *Türk Dünyası Dil ve Edebiyat Dergisi*, 56, 71-100.
DOI: 10.24155/tdk.2023.225

* Dr. Öğr. Üyesi, Artvin Çoruh Üniversitesi Eğitim Fakültesi Yabancı Diller Eğitimi Bölümü Yabancı Diller Ana Bilim Dalı, mehmetakkus@artvin.edu.tr, <https://orcid.org/0000-0002-9604-1418>.

** Dr., Artvin Çoruh Üniversitesi Fen-Edebiyat Fakültesi, gokcan@artvin.edu.tr, <https://orcid.org/0000-0002-6933-8494>.

değişkesi olan GAT arasında coğrafi yakınlığın sözlüksel etkisinin olup olmadığı da (alansal yakınsama) incelenmiştir. Çalışmada kullanılan veriler 2021 yılında Temmuz ve Eylül ayları arasında gerçekleştirilmiş olan alan araştırması sırasında İran'da Merkezî ve Kum vilayetlerinin Halaç Türkçesikonuşulan tüm yerleşim birimlerinden ve Kum vilayetinde Azerbaycan Türkçesi konuşulan ancak Halaç Türkçesi yerleşimlerle çevrili Couzeh köyünden derlenmiştir. Bu çalışma kapsamında aynı dil ekolojisi içerisinde konuşulmakta olan iki Türk dili değişkesinin sözlüksel uzaklığı algoritmik çerçevede ölçülerek sözcüklerin ses bilgisel uzaklıklarının söz konusu dil ekolojisi dâhilindeki oranları saptanmıştır. Elde edilmiş olan bulgular, Halaç Türkçesi ile GAT arasındaki sözlüksel uzaklığın paylaşılan ortak dil ekolojisine rağmen göreceli olarak anlamlı olduğuna işaret etmektedir.

Anahtar Sözcükler: Halaç Türkçesi, Levenshtein uzaklık algoritması, Dijkstra algoritması, ağız, Güney Azerbaycan Türkçesi.

Lexical Distance between the Khalaj Dialects and the Qom (Couzeh) Dialect of South Azerbaijani Turkic: An Amalgam of Levenshtein and Dijkstra Algorithms

Abstract

This study explores the lexical distance of two Turkic language varieties, namely Khalaj and South Azerbaijani Turkic, spoken in the Markazî and Qom provinces, based on the Leipzig-Jakarta core vocabulary list (Tadmor and Haspelmath, 2009). To this end, all Khalaj dialects spoken in Markazî and Qom provinces, as classified by Doerfer (1998), and the Couzeh variety of South Azerbaijani were compared in the present study. Levenshtein Distance Algorithm (LDA), frequently utilized in the relevant literature among quantitative dialectal measurement as a quantitative method, was used to determine the distance between the mentioned varieties. Using the emergent values, the minimum path between the Couzeh variety and Khalaj dialects was measured, using the Dijkstra Algorithm. The study also investigates whether there is any impact of geographical proximity (areal convergence) between Khalaj (Doerfer, 1971, 1987, 1988; Tekin, 1989) and South Azerbaijani. The data used in the study during the fieldwork carried out between July and September, 2021 were compiled from all Khalaj-speaking settlements in Markazî and Qom provinces in Iran, and from the village of Couzeh, where South Azerbaijani is spoken in the province of Qom. Within the scope of this study, the lexical distance of two Turkic varieties spoken in the same language habitat was measured in an algorithmic framework, and the ratios of the phonological distances of the lexical items were measured. The findings revealed that the lexical distance between Khalaj and Southern Azerbaijani is relatively significant despite the shared common language ecology.

Keywords: Khalaj, Levenshtein distance algorithm, Dijkstra algorithm, dialect, South Azerbaijani Turkic.

Giriş

1968 yılında Doerfer tarafından keşfine kadar Halaç Türkçesi, Doerfer'in öncülleri tarafından (Minorsky, 1940) yanlışlıkla İran'da konuşulan Güney Azerbaycan Türkçesinin bir varyantı olarak sınıflandırılmıştır. Birçokları ise *Divânu Lügati't-Türk*'te geçen Doerfer ve ark.na göre (1971) Argu dilinin günümüzdeki değişkesi olan Halaç Türkçesinin varlığından haberdar dahi değildi. Aynı dil ekolojisi içerisinde birbirine son derece yakın bir coğrafyada konuşulan Halaç Türkçesi ve Güney Azerbaycan Türkçesi, İran diyalektolojisi uzmanı Moghaddam'in (1939) ve Minorsky'nin (1940) sunmuş oldukları Azerbaycan Türkçesiyle karışık verilerle yayımlanmıştı. Doerfer'in 1968 yılında bu verilere erişimiyle Göttingen'de masa başında gerçekleştirmiş olduğu (yeniden) keşif ve sonrasında İran'ın Halaç Türkçesi konuşulan bölgelerinde ekibiyle gerçekleştirdiği Göttingen alan araştırması (1969-1971) neticesinde söz konusu bu değişkenin Azerbaycan Türkçesinden farklı bir değişke olduğu, derlemeler ve dil bilimsel incelemeler yoluyla sarıh bir şekilde ortaya konmuştur (Doerfer ve ark., 1971; Doerfer, 1987, 1988). Tekin (1989) bu doğrultuda "Türk Dil ve Diyalektlerinin Yeni Bir Tasnifi" başlıklı sınıflandırmasında Halaç Türkçesini *hadaq* grubu içerisinde tek başına bir grup oluşturacak şekilde tasnif etmiştir.

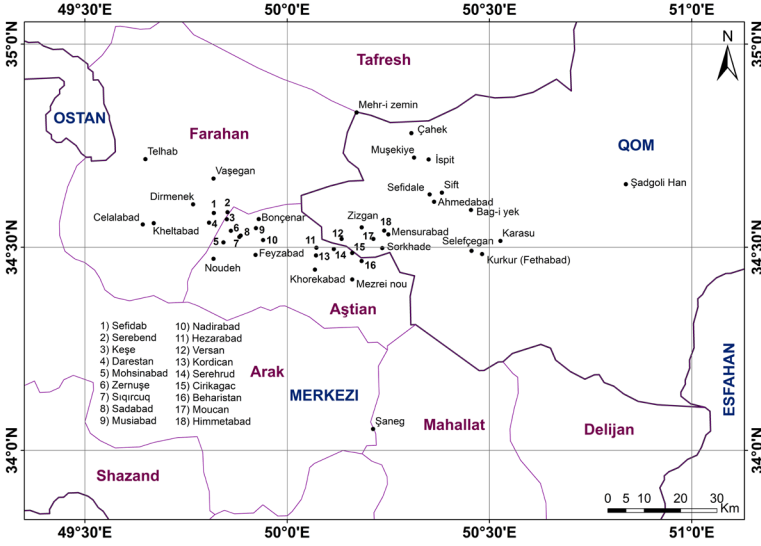
Bişim birim, ses birim ve söz varlığı açılarından sergilediği eskicil özellikler ile Türk dili değişkeleri arasında farklı bir yerde duran Halaç Türkçesinin içerisinde bulunduğu dil ekolojisi bağlamında diğer değişkelerle uzaklığı, güncel dil bilimsel ölçütlerle henüz tam anlamıyla incelenmemiştir. Halaç Türkçesinin ağızları arasındaki uzaklıkla ilgili Doerfer'in verilerine dayalı bir çalışma yayımlanmış olsa dahi İran'da konuşulan Türk dili değişkeleri arasındaki uzaklığı güncel konuşma verilerine dayalı olarak saptamayı amaçlayan bir çalışma ilgili alan yazında henüz neşredilmemiştir. Bu sebeple bu çalışma alan yazında söz konusu açığı sözlüksel bağlamda "hesaplamalı diyalektoloji" (İng. computational dialectology) yöntemlerinden Levenshtein uzaklık algoritması ve Dijkstra algoritmasını kullanarak kapatmayı hedeflemektedir. Bu çalışma, ilgili alan yazında Dijkstra algoritmasını "hesaplamalı diyalektoloji" alanında kullanan ilk çalışma olması hasebiyle Türk diyalektoloji alan yazınına bir katkı sunmayı amaçlamaktadır.

Halaç Türkçesi Dil Ekolojisi Görünümü

Orta İran'da Merkezî ve Kum vilayetlerinde konuşulmakta olan Halaç Türkçesinin içerisinde bulunduğu çok dilli dil ekolojisi dâhilinde İrani dillerin yanı sıra farklı Türk dili değişkeleri de konuşulmaktadır. Bu minvalde ilgili dil ekolojisinde İrani dillerden Farsça, Tatça, konargöçerlik dönem-

Halaç Türkçesinin ağızları ile Güney Azerbaycan Türkçesinin Kum (Couzeh) ağızı arasındaki sözlüksel uzaklık: Levenshtein ve Dijkstra algoritmaları amalgamı

lerinde Luri konuşulmaktayken Türk dili deęişiklerinden ise Şahseven Türkçesi, Merkezî ve Kum vilayetlerinde konuşulan Güney Azerbaycan Türkçesi deęişikleri, konargöçerlik dönemlerinde Kaşkay Türkçesi sayılabilir. Şekil 1, Halaç Türkçesinin ve Couzeh deęişkesinin İran’da konuşulduğu alanı işaretlemektedir.



Şekil 1: Merkezî ve Kum Vilayetlerinde Halaç Türkçesi Ağızları ve Güney Azerbaycan Türkçesinin Konuşulduğu Coğrafyanın Haritası

Dolayısıyla her ne kadar oldukça dar bir coğrafyada sınırlı sayıda dil topluluğu tarafından kullanılan bir dil olsa da Halaç Türkçesi birçok dil ve/veya dil deęişkesiyle etkileşim hâlinde görünmektedir.

Yukarıda dilsel manzarası tavsif edilen bölgeyi gösteren haritada işaretlenen çok dilli dil ekolojisi içerisinde yok olma tehlikesi altında bulunan ve bu durumdaki hemen her dilde tanıklandığı üzere varyantlaşmanın arttığı Halaç Türkçesi üzerindeki etkinin sözlüksel boyutları, bu çalışmanın temel konusunu teşkil etmektedir.

Bu bağlamda Halaç Türkçesi dil ekolojisi bünyesinde Halaç Türkçesini etkilediği düşünülen ilk dil, İran İslam Cumhuriyeti’nin resmî dili Farsça’nın bölgede konuşulan ağızlarıdır. Farsça, geniş Hint-Avrupa dil ailesinin İrani diller grubunda yer almakta ve Orta İran’dan Horasan’ın doğusuna kadar geniş bir coğrafyada ana dil olarak konuşulmaktadır. Deri ve Tacikçeyle göreceli karşılıklı anlaşılabilirliğe rağmen söz konusu diller, dil sınıflandırmalarında Farsçadan ayrı diller olarak tasnif edilmektedir (Ethnologue, 2022). Ana dil olarak konuşulmasının yanında Farsça, İran’da yaşayan

Farsçadan farklı ana dile sahip dil topluluklarınca (Azerbaycan Türkleri, Halaçlar, Türkmenler, Horasan Türkleri, Araplar, Beluçlar, Tatlar, Talişlar vb.) da ikinci dil olarak kullanılmaktadır. Halaç Türkçesinin konuşulduğu coğrafyada dengeli Halaç Türkçesi-Farsça iki dilliliğin yaygın olduğu bu minvalde ilgili alan yazında sıklıkla vurgulanmıştır (Doerfer ve ark., 1971).

Halaç Türkçesinin olduğu coğrafyada konuşulan diğer bir İrani dil Tatçadır (ISO 639-3 kodu: tks). Tatça, İran'ın kuzeydoğusunda konuşulan Hint-Avrupa dil ailesinin İrani dalına mensup bir dildir. Tatça, Halaç Türkçesi konuşulan coğrafyanın kuzeydoğusunda Amore gibi köylerde konuşulmaktadır.

Konargöçer dönemlerden Luri (ISO 639-3 kodu: lrc) dilli topluluklarla Halaçlar arasında iletişime dair bilgiler hem tarihî veriler hem hâlihazırda halkın toplumsal hafızasında yer alan hikâyelerle doğrulanmaktadır. Dil bilimsel veriler açısından ise Halaç Türkçesi Luri etkileşimini Kuribayashi (2012) Luri dilinden Halaç Türkçesine giren bir biçim birim örneğinde tanımlamıştır. Lurinin Halaç Türkçesinde biçim bilgisel düzeyde bir değişime neden olabilmesi için Halaç Türkçesi konuşurları ile Luri konuşurlarının tarihlerinin belli bir döneminde bir ortak yaşam sürmesi beklenmektedir. Luri, Windfuhr ve Perry (2009, s. 418) tarafından “Luri türevli” değişkeler olarak tanımlanmıştır. Luri ağızları, güneybatı İran'da tarihsel olarak Güney Erken Yeni Farsçadan evrilmiş birçok dil bilgisel yapıyı barındırdığı için “Farslaşmış” (İng. Perside) olarak tavsif edilmektedir. Bu noktada Luri diliyle ilgili alan yazında esas olarak Merkezî Luri, Bahtiyari, Boyer Ahmedi ve Mamasani-Kuhgeluye ağızları tasnif edilmektedir. Anonby (2012), Luri konuşur sayısının 4 milyon ila 5 milyon arasında olduğunu belirtmektedir.

Güney Azerbaycan Türkçesi (ISO 639-3 kodu: azb) genel olarak Halaç Türkçesi konuşulan dil coğrafyasının kuzeyine tekabül eden köylere yakın bölgelerde sayıca büyük bir Türk dilli nüfus tarafından konuşulmaktadır. Halaç coğrafyası içerisinde Telkhab, Himmetabad, Esfid, Muşekiye gibi köylerde Azerbaycan Türkçesi konuşurlarının sayısının yıldan yıla arttığı kaydedilmektedir. Bu nedenle özellikle kuzeyde Azerbaycan Türkçesi konuşulan bölgeye yakın Halaç köylerinde Oğuzcanın göreceli olarak etkili olduğu belirtilebilir.

Şahseven Türkçesi, İran'ın kuzeybatı bölgelerinde konuşulan ve Güney Azerbaycan Türkçesinin bir kolu olarak kaydedilen ve zaman zaman Türkiye Türkçesinin doğu grubu ağızlarına yakınsayan özelliklere sahip bir değişke olarak tanımlanır (Çam, 2021). Bu nedenle ISO kodu Güney Azerbaycan Türkçesi dâhilinde değerlendirilmektedir. Erdebil ve Zencan bölgesi Şahseven ağızları üzerine çalışmalar yapılmışsa da Merkezî ve Kum vilayetlerinde Halaç Türkçesi konuşulan bölgeye yakın kullanılan Şahseven ağızları üzerine henüz kapsamlı bir karşılaştırmalı çalışma yayımlanmamıştır.

Diyalektometri Çalışmaları

Ekseriyetle tarihsel dil biliminde kullanılmış olan leksikoistatistik (=glot-tochronology) çalışmalarıyla art zamanlı başlayan nicel ölçüm çalışmaları (Gudschinsky, 1956), 21. yüzyıla gelindiğinde makine öğrenimi (İng. machine learning) ve yapay zekâ (İng. artificial intelligence) çalışmalarındaki algoritmik gelişmelere uygun olarak gelişmiştir.

Ağızların birbirleriyle ilişkilerini tek bir üst yapıya bağlamadan saptamak için “diyalektometri” adı verilen bir yöntem kullanılmaktadır (Szmrecsanyi, 2013). İlk defa Fransız lehçe bilim çalışmalarının bir hediyesi olarak Séguy (1973, s. 1) tarafından alan yazına kazandırılmış olan “dialectométrie” kavramı, derlenmiş verilerin varyasyon yoğunluğuna bir çözüm olarak nicelleştirilmesi amacıyla kullanılmıştır. Hesaplamalı diyalektometri yöntemlerinden “kategorik karşılaştırmalar”dan (İng. categorical comparisons) biri “göreceli özdeşlik değeri”dir (İng. Relative Identity Value). Séguy (1973) ve ekibi iki komşu ağız arasındaki dil bilimsel mesafeyi, üzerinde anlaşamadıkları “uzak(lık)” öğelerinin sayısı temelinde saptamışken Goebel (1982), farklılıklar yerine benzerlikleri ölçmüş ve buna “Göreceli Özdeşlik Değeri” adını vermiştir. Benzerlik ve uzaklık kategorileri temelinde geliştirilmiş olan bu göreceli değer yaklaşımı sonradan geliştirilmiş olan hesaplamalı diyalektometriye yön vermiştir.

Regensburg-Salzburg Diyalektometri (RS-DM) ekibi bilgisayar yöntemlerini kullanarak ses bilgisi temelinde haritalandırılmış ağızlararası benzerlik temelli çalışmalarıyla diyalektometri çalışmalarına kuramsal ve uygulamalı katkılar sağlamıştır (bkz. Goebel, 2017). Bu çalışmalar sonucunda göreceli özdeşlik değerinin yanı sıra belli birtakım dilsel özelliklerin “daha önemli”, diğerlerinin göreceli olarak “daha az önemli” olduğunu savlayan “Ağırlıklı Özdeşlik Değeri” (İng. Weighed Identity Value) mefhumunu da alan yazına tanıtmıştır.

Bir diğer hesaplamalı diyalektometri yöntemi “sıklık temelli yöntemler” (İng. frequency-based method) olarak gruplandırılan sessel ve ses bilimsel özelliklerin kullanım sıklığını dikkate alan yöntemlerdir (Heeringa ve Prokić, 2017). Sessel özellikler yanında Szmrecsanyi (2008) Freiburg İngilizce Derlemi (FRED) kullanarak İrlanda dışındaki Büyük Britanya İngilizcesinde sözdizimsel özellik sıklıklarını incelemiş ve Öklid uzaklığını kullanarak aralarındaki farkları saptamayı amaçlamıştır.

Diyalektolojinin Avrupa ekolünce Romen dillerinin (Goebel, 1982) veya Felemenkçe ağızlarının (Nerbonne ve Heeringa, 2010) akrabalık ilişkilerinin saptanmasında nicel yöntemler uyarlanmaya ve geliştirilmiş olan algoritmalarla yararlanılmaya başlanmıştır. Toplum dilbilimsel etkenleri görmezden geldiğini savunanlar tarafından bilimsel eleştiriye tabi tutulan söz konusu al-

goritma temelli nicel diyalektometri yöntemleri ağız ve lehçeler arası uzaklık ölçümlerinde günümüzde sıklıkla kullanılmaya devam etmektedir.

Bu ölçüm tekniklerine genel olarak *Dizge Düzen Uzaklığı* (İng. String Edit Distance) adı verilmektedir. Bu teknikler arasında ağızlararası uzaklığın ölçümünde ilgili alan yazında en çok kullanılanı *Levenshtein Uzaklık Algoritması* ve *n-Gram Ağırlıkları* (İng. n-Gram Weights) olmuştur. Bu ölçüm teknikleri göreceli (İng. relative) ve mutlak (İng. absolute) uzaklıkları ölçüp ölçmeme konusunda farklı paradigmaları göstermektedir.

Alan yazında Türk dili değişkeleri içerisinde sadece Halaç Türkçesi ağızları (güney, kuzey ve kuzeydoğu) arasındaki uzaklığın saptandığı bir çalışma olduğu görülmektedir. Söz konusu çalışmada kuzey ağızlarının birbirlerine sözlüksel olarak yakınsadığını güney ağızı ile ise göreceli bir uzaklığın olduğunu ileri sürülmüştür. Bu ağızların birbirleriyle uzaklık ilişkisini diyagram ve algoritma temelinde gösterme konusundaki eksiklik ise eldeki bu çalışmanın yapılmasını gerekli kılmıştır.

Bu yönüyle bu çalışma üç ve üçten fazla ağızın ölçüldüğü durumlarda en kısa yoldan bu ağızlar arasında merkeze alınan ağızla ilişkileri alan yazına tanıtacak Dijkstra algoritmasının diyalektoloji çalışmalarında kullanımının uygunluğunu tartışmaya açması açısından da özgündür.

Veri Toplama ve Analiz Yöntemi

Bu çalışma kapsamında İran'ın Kum vilayetinde birbirine komşu coğrafyada konuşulan bir Güney Azerbaycan Türkçesi değişkesi (Couzeh) ile Halaç Türkçesi arasındaki sözlüksel uzaklığın saptanması amaçlanmaktadır.

Araştırma Bağlamı: Alan Araştırması

Bu çalışma kapsamında güvenilir (İng. trustworthy) sözlü veri toplamak maksadıyla İran İslam Cumhuriyetinin Merkezî ve Kum sancaklarına Temmuz-Eylül 2021'de araştırma gezisi düzenlenmiştir.

Alan araştırması sırasında veri derlenmiş olan Halaç Türkçesi ve Güney Azerbaycan Türkçesi konuşulan yerleşim birimlerinin bağlı bulunduğu sancak, vilayet ve ilçenin adları ile birlikte bu birimlerin coğrafi koordinatları aşağıdaki tabloda sunulmuştur:

Tablo 1: Veri Derlenmiş Olan Halaç Köyleri

	Sancak	Vilayet	Ağız	Enlem	Boylam
<i>Mensurabad</i>	Kum	Merkez	Merkez doğu	34.53156	50.25009
<i>Behāristan</i>	Merkezî	Āştiyan	Merkez	34.4662	50.18412
<i>Şaneg</i>	Merkezî	Erak	Güney	34.05283	50.21231
<i>Mehr-i Zemin</i>	Kum	Caferabad	Kuzey	34.83165	50.17165

Kheltabad	Merkezî	Ferahan	Batı	34.55924	49.67021
Karasu	Kum	Merkez	Ana ağız	34.51557	50.52709
Muşekiye	Kum	Caferabad	Kuzey doğu	34.72077	50.31379
Couzeh	Kum	Merkez	GAT Kum	34.53840	50.21099

Tablo 1, Doerfer'in (1998) belirlemiş olduğu 7 Halaç Türkçesi ağız bölgesinden bu çalışmada verisi kullanılan 7 Halaç yerleşim birimini ve karşılaştırma için aynı dil ekolojisi içerisinde konuşulan GAT değişkisi Couzeh ağızını göstermektedir. Bunlar sırasıyla Mensurabad (Merkez doğu ağız bölgesi), Behāristan (Merkez ağız bölgesi), Şaneg (Güney ağız bölgesi), Mehr-i Zemin (Kuzey ağız bölgesi), Kheltabad (Batı ağız bölgesi), Karasu (Ana ağız bölgesi), Muşekiye (Kuzey doğu ağız bölgesi) ve Couzeh (GAT).

Kum ve Merkezî sancaklarından çalışmaya katılan kaynak kişilerin demografik bilgileri aşağıdaki tabloda sunulmuştur:

Tablo 2: Veri Derlenmiş Olan Halaç Türkçesi Konuşurları

Kaynak Kişi	Sancak	Köy	Yaş	Cinsiyet	Eğitim
Kazım Ali Mededi	Kum	Mensurabad	44	Erkek	Ortaöğretim
Tavus Mollayi	Merkezî	Behāristan	85	Kadın	Eğitimsiz
Emir Hatibi	Merkezî	Şaneg	29	Erkek	İlköğretim
Huseyn Şanegi	Merkezî	Ferahan	45	Erkek	İlköğretim
Hasan Recebali	Kum	Mehr-i Zemin	60	Erkek	İlköğretim
Emir Hac'hamidi	Kum	Karasu	66	Erkek	İlköğretim
Kadir Ali Şerefzade	Kum	Muşekiye	51	Erkek	Ortaöğretim
Sadeq Celalabadi	Merkezî	Heltabad	73	Erkek	İlköğretim
Ebulfezl Heyderpenah	Kum	Couzeh	83	Erkek	İlköğretim

Tablo 2'deki verilerin işaret ettiği üzere çalışmaya katılan kaynak kişilerin büyük çoğunluğu erkeklerden (N=8) müteşekkil iken İran'ın kendine has toplumsal normlarının da etkisiyle sadece bir kadın (N=1) çalışmaya katılabiliştir. Tüm katılımcılar içerisinde en genç kaynak kişinin yaşı 29 olarak belirlenmiştir. Çalışmanın yapıldığı zaman kayıt alınan en yaşlı kaynak kişi ise 85 yaşından gün almaktaydı. Tüm katılımcıların ortalama yaşı ise 59,55 olarak hesaplanmıştır. Katılımcıların eğitim düzeylerine gelince, kaynak kişilerden kadın olanın eğitim almadığı tespit edilmiştir. Kaynak

kişilerin kahir ekseriyeti ya ilköğretim (N=6) veyahut ortaöğretim (N=2) seviyesinde eğitim aldıklarını belirtmişlerdir. Katılımcıların isim bilgilerinin yayımlanması için gerekli onamlar ve Etik Kurul izinleri alınmıştır.

Veri Toplama Aracı: Leipzig-Jakarta Sözcük Listesi

Ağızlararası uzaklığı ölçmek ve sözlüksel öğelerin nicel bir çözümlemesini sunabilmek adına “ödünclemeye direnç”, “evrensellik”, “sadelik” ve “istikrar” gibi değişkenler dikkate alınarak Tadmor ve Haspelmath tarafından geliştirilmiş olan ve “Leipzig-Jakarta Listesi” (EK 1) olarak tesmim olunan 100 maddelik temel bir sözcük listesi oluşturulmuştur (bkz. Tadmor, 2009).

Bu listede 49 ad, 17 önad (sıfat), 24 eylem, 2 belirteç (zarf), 6 adıl ve 1 edat bulunmaktadır. Bu çalışma kapsamında tüm sözlüksel öğeler Levenshtein algoritması temelli olarak -makalede sayfa kısıtlamaları nedeniyle- incelenemeyeceğinden her bir dil bilgisi kategorisinden bir örnek çözümleme sunulmuştur.

Bu liste alan gezilerine başlamadan önce Behāristan köyünden kılavuz kişiler Omid (Afshin) Arabgol ve Azer Arabgol tarafından gözden geçirilmiş, değerlendirmelerden sonra sözlük maddeleri Farsçaya çevrilmiştir. Çeviriler Farsça yeterliği olan kılavuz kişilerce kontrol edilmiştir. Gerekli düzenlemeler yapıldıktan sonra liste, alt alan adlarının Loanword Typology (LWT) listesinde yer alış sırasına göre 6 gruba ayrılmıştır. Bu doğrultuda, fiziksel dünya, akrabalık, hayvan ve vücut bölümleri gibi somut sözcük öğeleri konuşurlarca kolayca görselleştirebilmesi amacıyla PowerPoint sunumu biçiminde hazırlanmıştır. Sözcük listelerinin ortaya çıkartılması için görsel ipuçları yanında soyut mefhumlar için Farsça karşılıkları sözlü ipucu olarak kullanılmıştır.

Bu amaçla bölgeden derlenen veriler LUA aracılığıyla incelenmiştir. Güney Azerbaycan Türkçesi değişkesi ile Halaç Türkçesi ağızları arasındaki LUA ile yapılan ölçümler birer kenar yükü ve ilgili değişke ve ağızlar birer köşe kabul edilerek Dijkstra algoritması yardımıyla kaynak olarak alınan bir ağızdan diğerlerine minimum yollar elde edilerek bu ağızlar arası bir ağaç oluşturulması hedeflenmiştir.

Veri Çözümleme Yöntemi I: Levenshtein Uzaklık Algoritması

Dilbilimde özellikle karşılıklı anlaşılabilirlik çalışmaları ve algısal çok dilcilik temelinde de kullanılan bir metrik olan LUA, dil yakınlığı (İng. language proximity) veya iki dil değişkesi ve ağız arasındaki yakınlık/uzaklık durumunu ölçmek için kullanılmaktadır. Bu algoritma, ağız varyantlaşmasının saptanmasında dünya üzerinde birçok dile başarıyla uygulanmıştır:

Hollandaca (Heeringa, 2004; Wieling vd. 2007), Sardunyaca (Bolognesi ve Heeringa, 2002), Norveççe (Gooskens ve Heeringa 2004), Almanca (Nerbonne ve Siedle, 2005) ve Bulgarca (Osenova vd. 2009). Geçmişe nazaran bilgiye çok daha kolay ulaşılan 21. yüzyılda bilgisayar ve genel ağ teknolojisinin dil ve dil verilerinin çözümlenip yorumlanmasına katkısı yadsınamaz bir gerçek olarak karşımıza çıkmaktadır. Bilgisayar ve genel ağ teknolojisinin temelinde yatan gerçek ise algoritmalarıdır. Anlaşılması güç ve zaman alan birçok problem, çok sayıda koşulun ve sorunun iç içe yer aldığı birçok problem algoritma şemalarına aktarılarak çözümlenebilir hale gelmektedir. Özellikle 21. yüzyılın bir konusu olan algoritmalar geçmişten günümüze daha net ve daha hızlı sonuçlar vermesi adına sürekli geliştirilmektedir. Bu bağlamda sayısallaştırılmış olan dil verilerinin algoritmik ölçümlerinin bilimsel nesnellik açısından sağladığı katkı göz ardı edilemeyecek denli önemli görülmektedir. Özellikle diyalektolojinin Avrupa ekolünün alanda uygulamaya geçirdiği ve geliştirmeye devam ettiği algoritma ve dizge düzen uzaklığı temelli çalışmaları bu alanda öncü rol oynamıştır.

Çalışmanın bu kısmında verili iki dizgi arasında, bir dizgenin diğer dizgeye dönüştürülmesi için minimum işlem sayısını veren LUA algoritması incelenmektedir. Algoritma adını 1965 yılında verili iki dizge arasındaki mesafeyi ölçen Sovyet matematikçi Vladimir Levenshtein'dan almıştır. Bir dizgeyi diğer bir dizgeye dönüştürmek için yapılabilecek işlemler ekleme, silme ve değiştirmedir. LUA algoritmasında birden fazla yöntem vardır. Bunlar arasında en sık kullanılanı dinamik programlama yöntemidir. Dinamik programlama yöntemi iki aşamadan oluşmaktadır. İlk aşamada verilen iki dizge arasında değerler matrisi oluşturulur. Bu adımda bir dizgenin diğer dizgeye benzemesi için kaç değişikliğe ihtiyaç duyulduğuna karar verilir. İkinci aşamada ise bu değişikliklerin nasıl tanımlanacağına karar verilir. Diğer bir deyişle ekleme, silme ve değiştirme işlemleri tanımlanır. Bu tanımlamada ilk sözcüğün ve ikinci sözcüğün son seslerinin LUA ölçümlerinden köşegen üzerinde geri izleme yapılır. Köşegen üzerinde değişmeyen ölçümler bir işlem yapılmaması gerektiği anlamına gelir. Köşegen üzerinde değişen ölçümler değiştirme işlemine, sola doğru azalan ölçümler silme işlemine ve yukarı yönlü azalan ölçümler ise ekleme işlemine karşılık gelir.

Algoritmanın uygulanmasında dizgeler çapraz olarak bir matrise yerleştirilir. Dizgelerin başına boş satır ve sütun açılır. Satır olarak yazılan dizgenin seslerine 1'den başlayarak numaralar verilir. Birinci sözcüğün ilk sesinden son sesine kadar her ses ikinci sözcüğün sesleri ile eşleştirilir. Eşleşmede eğer sesler aynı ise çaprazda bulunan sayının üstündeki, solundaki ya da sol üstteki sayıdan küçük olanı değer olarak alır, aksi takdirde

sayının üstündeki, solundaki ya da sol üstündeki sayıların değeri 1 artar bu durumda küçük olanı değer olarak alır. Örneğin a dizgesi i sayıda ses ve b dizgesi j sayıda ses içersin. a ve b dizgeleri arasındaki matrisin (i, j) elemanının değerini veren LUA ölçümü aşağıdaki gibi verilebilir:

$$\text{lev}_{a,b}(i, j) = \begin{cases} \max(i, j) & \text{If } \min(i, j) = 0 \\ \min \begin{cases} \text{lev}_{a,b}(i-1, j) + 1 \\ \text{lev}_{a,b}(i, j-1) + 1 \\ \text{lev}_{a,b}(i-1, j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{Otherwise} \end{cases}$$

Genel olarak iki dizge arasındaki LUA ölçümü Şekil 2’de gösterilen algoritma ile verilir:

```

Function LEVDIST ( $a, b$ ) is
   $n \leftarrow a.size()$ ;
   $m \leftarrow b.size()$ ;
   $distance[i, 0] \leftarrow i \forall i \in 0..n$ ;
   $distance[0, j] \leftarrow j \forall j \in 0..m$ ;
  for  $j \leftarrow 1$  to  $m$  do
    for  $i \leftarrow 1$  to  $n$  do
       $indicator \leftarrow a[i] \neq b[j] ? 1 : 0$ ;
       $distance[i, j] \leftarrow \text{Minimum} \left( \begin{array}{l} distance[i-1, j] + 1, \quad // \text{deletion} \\ distance[i, j-1] + 1, \quad // \text{insertion} \\ distance[i-1, j-1] + indicator, \quad // \text{substitution} \end{array} \right)$ ;
    end
  end return  $distance[n, m]$ ;
end

```

Şekil 2: Levenshtein Uzaklık Algoritması

Şekil 2’de yazılmış olan algoritmada görüldüğü üzere, Levenshtein uzaklığının birkaç alt ve üst sınırı vardır. Bu sınırlar, en az iki dizginin farkı olarak bulunmaktadır. En fazla ise uzun dizginin uzunluğuna denk gelmektedir. Sadece iki dizgi birbirine eşitse sayısal değer sıfır olarak saptanmaktadır. Aynı dizgiler birbirine denkse *ekleme*, *silme* ve *değiştirme* işlemlerinden herhangi biri işlem değeri olarak kabul edilmemektedir. Ekleme, silme veya değiştirme gibi işlemlerin gerçekleştirildiği her bir adımda ise uzaklığın sayısal değeri 1 artmaktadır. Basit algoritmik işlemler açısından örneğin “[akar]”, “[eker]” dizgeleri arasındaki uzaklık 2’dir çünkü ilk ve ikinci /a/ sesleri /e/ sesleri ile değişmiştir. “[akar]”, “[akar]” dizgeleri arasındaki uzaklık ise 0’dır çünkü birbirlerinin aynı ses ve biçim özelliklerine sahip bu iki sözcüğün birbirlerine benzeşmesi için herhangi bir işlem yapılması gerekmemektedir. Dolayısıyla LUA algoritmasına göre bu iki dizge arasında herhangi bir uzaklık bulunmamaktadır. Bununla birlikte, aşağıda örneklendirildiği gibi, birtakım sözcükler arasındaki işlem

farkları daha karmaşık olabilmektedir. Levenshtein algoritması, değişim değerini saptamak üzere iki boyutlu bir dizi temelinde sözcüklerin değişik olan sesleri için sayısal değerine gitmektedir. Bu bağlamda örnek olması için, Tablo 3, Halaç Türkçesinin farklı ağızlarında “yumurta” anlamına gelen [jumurta] dizgisi ile [numurqa] dizgisi arasındaki uzaklığın LUA temelinde çözümleme yolunu sunmaktadır.

Tablo 3 “yumurta” ve “numurqa” sözcükleri arasındaki mesafeyi LUA ölçümüne göre işlemsel olarak sunmaktadır:

Tablo 3: “yumurta” ve “numurqa” Sözcüklerinin Matrisinin Oluşturulması

	0	Y	U	M	U	R	T	A
0	0	1	2	3	4	5	6	7
N	1	1	2	3	4	5	6	7
U	2	2	1	2	2	3	4	5
M	3	3	2	1	2	3	4	5
U	4	4	2	2	1	2	3	4
R	5	5	3	3	2	1	2	3
Q	6	6	4	4	3	2	2	3
A	7	7	5	5	4	3	3	2

Tablo 3’teki işlemin işaret ettiği üzere, “yumurta” ve “numurqa” dizgeleri arasındaki LUA ölçümü 2’dir.

Dinamik programlamada ikinci işlem, “yumurta” dizgesini “numurqa” dizgesine dönüştürmek için hangi işlemlerin yapılacağına karar vermektir. Bu sözcüklerin son seslerinin LUA ölçümünden köşegen üzerinde geri izleme yapılırsa iki defa ölçümde azalma görülür. Bunlar değiştirme işlemleri olarak tanımlanır. Şekil 3 ise, Güney Azerbaycan Türkçesinin Couzeh ağızı ile Halaç ağızları arasındaki sırasıyla Türkiye Türkçesinde

“bir” sayı önadı (GAT Couzeh: /bi:/; Halaç Türkçesi: /bi:/),

“ne” soru adlı (GAT Couzeh: /næ/; Halaç Türkçesi: /nɛsɛ/),

“ateş” adı (GAT Couzeh: /od/; Halaç Türkçesi: /huot/),

“düşmek” (GAT Couzeh: /diʃmæk/; Halaç Türkçesi: /tiʃmæk/) eylemi,

“ben” adlı (GAT Couzeh: /mæn/; Halaç Türkçesi: /mæn/) ve

“dün” (GAT Couzeh: /dɯnɛn/; Halaç Türkçesi: /ængir/) zaman belirteci anlamına gelen sözlüksel öğelerin program aracılığıyla işlem çözümleme sonuçlarını göstermektedir.

```

C:\Users\Hafiz\Documents\LevenshteinMesafe\LevenshteinMesafe\bin\Debug\LevenshteinMesafe.exe
01: vs 01:
Mesafe:0
0 1 2 3
1 0 1 2
2 1 0 1
3 2 1 0

02: vs 02:
Mesafe:2
0 1 2 3 4
1 0 1 2 3
2 1 0 1 2

03: vs 03:
Mesafe:3
0 1 2
1 1 2
2 2 2
3 2 3
4 3 3

04: vs 04:
Mesafe:1
0 1 2 3 4 5 6
1 0 1 2 3 4 5
2 1 1 1 2 3 4 5
3 2 2 2 1 2 3 4
4 3 3 3 2 1 2 3
5 4 4 4 3 2 1 2
6 5 5 4 3 2 1

05: vs 05:
Mesafe:0
0 1 2 3
1 0 1 2
2 1 0 1
3 2 1 0

06: vs 06:
Mesafe:5
0 1 2 3 4 5
1 1 2 3 3 4
2 2 3 3 3 4
3 3 4 4 4 4
4 4 4 4 4 4
5 5 5 5 5 5

```

Şekil 3: Levenshtein Uzaklık Algoritması Program Çözümleme Sonuçları

Veri Çözümleme Yöntemi II: Dijkstra Algoritması

Bilgisayar ve genel ağ teknolojisinin temelinde yatan algoritmalar in- sanoğlunun yaşamını kolaylaştırmakta ve teknolojik gelişmeler doğrultu- unda algoritmaları geliştirme süreçleri devam etmektedir. Algoritmalar sonuç değil sonuca giden yolu veren şemalardır. Kullanılacak algoritmanın seçimi problemin türüne göre farklılık gösterir. 1959 yılında W. E. Dijk- stra tarafından önerilen algoritmada verili köşeler arasında keyfi iki köşe arasındaki minimum uzunluk ya da belirli iki köşe arasındaki minimum uzunluğun bulunması amaçlanır. Dijkstra (1959) ve Whiting ve Hillier (1960) tarafından tanımlanan algoritmalar bu amaçla tanımlanan en verimli algoritmalar (Dreyfus, 1969). Bu algoritmada köşeler ve köşeler arası kenarlar vardır, bu nedenle doğal olarak graflar üzerinde çalışılır. Kenarlar yönlü ya da yönsüz olabilir. Kenarlar pozitif yüklü olmalıdır. Negatif yüklü olması durumunda Dijkstra algoritması çalışmaz. Dijkstra algoritmasında köşeler geçici ve kalıcı değerler alır. Minty, Pollack ve Wiebenson tarafın- dan Dijkstra algoritmasının uygulanmasında köşelerin kalıcı değerlerinin elde edilmesi için tanımlanmıştır (Dreyfus, 1969).

Dijkstra algoritmasında ilk önce kaynak bir köşe seçilir. Kaynak köşe- nin değeri 0 diğer köşelerin değerleri belirsiz olduğu için ∞ (sonsuz) deęerini alır. Burada kaynak köşenin değeri kalıcı diğer köşelerin ∞ değerleri geçicidir. Kaynak köşeden grafin yönlenişine göre ulaşılacak olan ilk köşelere kenarlardaki yükler değer olarak atanır. Bu köşelere daha önce atanan ∞ değerleri ile yeni atanan değerler arasında kıyaslama yapılır, her

türlü yeni değerler ∞ 'dan küçük olacağından yeni değerler olarak atanır. Grafın yönelişine göre tüm graf dolaşılır ve verili iki köşe ya da kaynak köşeden diğer köşelere minimum uzunluklar hesaplanarak minimum yollar elde edilir.

Dijkstra algoritmasının akış şeması aşağıdaki gibi verilebilir:

a ve b birer köşe olmak üzere ikisi arasındaki uzaklık $x(a,b)$ olsun. a Köşesinin değeri $v(a)$ olmak üzere, Dijkstra algoritması köşeleri geçici ve kalıcı olarak etiketlediğinden a köşesinin etiket değerleri

$$l(a) = \begin{cases} 1, & \text{Kalıcı Etiket} \\ 0, & \text{Geçici Etiket} \end{cases}$$

olarak tanımlansın. Kaynak köşe s olmak üzere, $v(a)$; $s-a$ uzunluğunun en küçük değeri olarak tanımlanabilir. Eğer burada a 'nın öncülü varsa

$$\tau(a) = \begin{cases} 1, & \text{Varsa} \\ 0, & \text{Yoksa} \end{cases}$$

olarak alınsın.

1. Adım: Kaynak köşe için $v(s) \rightarrow 0$ ve $l(s)=1$ değerleri elde edilir.

Grafın diğer tüm a köşeleri için $v(a) \rightarrow \infty$, $l(a) = 0$, $\tau(a) \rightarrow 0$ olarak bulunur. d grafta bir köşe olmak üzere $s \rightarrow d$ olduğunu kabul edelim.

2. Adım: (d,a) kenarı için $l(a)=0$ ve $v(a) > v(d) + x(d,a)$ olarak alınsın.

Buradan $v(d) + x(d,a) \rightarrow v(a)$, $\tau(a) \rightarrow d$ elde edilir.

3. Adım: $l(a^*)=0$, $v(a^*) < \infty$ ve $v(a^*) = \min_{l(a)=0} \{v(a)\}$ koşulunu sağlayan a^* bulalım. Böylece $l(a^*) \rightarrow 1$ ve $a^* \rightarrow d$ elde edilir.

Bu adımda verildiği şekilde a^* köşesinin olmaması s 'den a 'ya bir yönlü yol olmadığı anlamına gelir ve algoritmanın çalışması durdurulur.

4. Adım: h Bir köşe olmak üzere, eğer $d \neq h$ ise, 2. adıma gidilir.

5. Adım: Dur. (Ruohonen, 2008).

Graflar ve Uygulamaları Hakkında Bazı Bilgiler

Graflar köşeler ve kenarlardan oluşur ve Graf Teorisi'nin temelini oluşturur. Graf Teorisi'nin oluşturulmasında esas olarak birkaç varsayım ileri sürülür. Bunlardan ilki Platon'a dayandırılır. Platon'un düzgün cisimlerinin açılması sonucu köşeler ve köşeler arası kenarlar oluşur.

İkinci varsayım ise Euler'e dayandırılır. Kaliningrad kasabasından geçen Königsberg nehri üzerinde kurulan 7 (yedi) köprüden geçme olayına dayalı bir oyun, 1783 yılında matematikçi Euler tarafından "Königsberg'in 7 (yedi) köprüsü" isimli çalışmayla matematik literatürüne kazandırılmıştır. Buradaki oyun, "Köprü'nün bir köşesinden başlayarak ve geçilen bir

köprüden tekrar geçilmeden aynı köşeye gelinebilir mi?” sorusuna yanıt bulmaya dayanmaktadır. Burada köprülerin uçları köşeler ve köprüler ile köprüler arası mesafeler kenarlar olarak temel alınmıştır.

Graf Teorisi’nde kullanılan bazı tanımlar aşağıdaki gibi özetlenebilir:

Bir kenarın iki ucunda yer alan noktalara *köşe*, iki köşe arasında yer alan uzunluğa ise *kenar* denir. Bir grafta eğer kenarlar yön bilgisi içeriyorsa *yönlü graf*, içermiyorsa *yönsüz graf* olarak adlandırılır. Bir grafta kenarların hepsi yön bilgisi içerir ya da hiçbirisi yön bilgisi içermez. Bir grafta, bir köşeden diğerine gidilirken izlenen kenarların toplamı yolu verir ve bir köşeden diğerine gidilirken minimum uzunlukların toplamı minimum yolu verir. Grafta mevcut bir köşeden başlayıp yine aynı köşeye dönen ve bir köşeden iki kez geçmeyen grafa *döngü* denir. Döngü içermeyen bir grafa ise *ağaç* denir.

Örnek. Tablo 1’de verilen Halaç Türkçesi ve Güney Azerbaycan Türkçesi ağızları arasındaki diyalektik farklar yüz (100) sözcük üzerinden LUA algoritmasına göre ölçülmüştür. İki ağız arasındaki fark ise

$$d(\text{Ağız 1}, \text{Ağız 2}) = \frac{\sum \text{lev}_{a,b}(i,j)}{100}$$

aritmetik ortalaması ile hesaplanmıştır.

Veri Analizi: Levenshtein ve Dijkstra Algoritması Çözümlenmeleri

Buradan ilgili ağızlar arasındaki LUA algoritmasına göre ölçülen ağız farkları Tablo 4’te gösterildiği şekilde hesaplanmıştır:

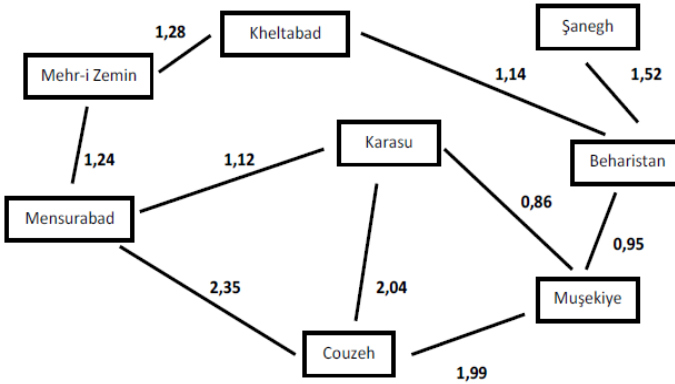
Tablo 4: Halaç Türkçesi Ağızları Arasındaki LUA Algoritmasına Göre Ölçülen Ağız Farkları

	Khelta- bad	Kara- su	Beha- ristan	Mensu- rabad	Mehr-i Zemin	Muşe- kiye	Şa- neg	Cou- zeh
Couzeh	2,54	2,04	2,14	2,35	2,48	1,99	2,38	0
Kheltabad	0	1,21	1,14	1,33	1,28	1,15	1,47	2,54
Karasu	1,21	0	1	1,12	1,36	0,86	1,28	2,04
Beharistan	1,14	1	0	0,94	0,93	0,95	1,52	2,14
Mensurabad	1,33	1,12	0,94	0	1,24	1	1,35	2,35
Mehr-i Zemin	1,28	1,36	0,93	1,24	0	0,99	1,74	2,48
Muşekiye	1,15	0,86	0,95	1	0,99	0	1,32	1,99
Şaneg	1,47	1,28	1,52	1,35	1,74	1,32	0	2,38

Tablodaki veriler ışığında, Levenshtein algoritmasına göre Couzeh ağızına sözlüksel uzaklığın en az olduğu Halaç Türkçesi ağızının Muşekiye

(kuzey doğu) ağızı (LUA=1,99) olduğu görülmektedir. Sonrasında Couzeh ağızına en çok yakınsayan ağız LUA=2,04 sonucuyla Karasu (ana ağız) ağızı olarak saptanmıştır. Beharistan (merkezi ağız) ağızı ile Couzeh ağızı uzaklık sonucu 2,14 olarak belirlenmiştir. Mensurabad (merkez doğu) ağızı (LUA=2,35) ve Şaneg (güney) ağızı (LUA=2,38) birbirine yakın değerlerle sözlüksel uzaklık bağlamında Couzeh ağızı ile göreceli olarak uzaklaşırken Mehr-i Zemin (kuzey) (LUA=2,48) ve Kheltabad (batı) (LUA=2,54) ağızları ise en uzak ağızlar olarak hesaplanmıştır. Levenshtein algoritması iki dizge arasındaki farkı ve/ya benzerliği incelediği için bu çözümlemede çok boyutlu çözümleme ile tüm Halaç Türkçesi ağızlarının Couzeh ağızı ile ilişkisini tespit etmek üzere söz konusu veriler Dijkstra algoritmasıyla da incelenmiş ve graf örneği çıkartılmıştır. Burada vurgulanması gereken nokta Levenshtein algoritması iki ağız arasındaki uzaklığı ortaya çıkartırken Dijkstra algoritması ise GAT değişkesi Couzeh merkezinde yedi Halaç Türkçesi ağızı arasındaki ilişkiyi incelemektedir.

Bu bağlamda öncelikle bu çalışmada, Tablo 4'te sunulan verilerin Dijkstra algoritmasına uyarlanması için ağızlar *köşe*, ağızlar arası farklar ise iki köşe arası *kenarlar* olarak alınmıştır. Bu minvalde Dijkstra algoritması temelinde Tablo 4'te sunulmuş olan Halaç Türkçesi ağızları ve GAT ağızı Couzeh arasında toplam 28 yol vardır. Bu çalışmada bu yollardan sadece 10 tanesi kullanılarak Şekil 4'te örneklendirilen bir *graf* örneği oluşturulabilir. Bu oluşturulan grafta ağızların konuşulduğu konumlar -bu çalışmanın odağı coğrafi uzaklığın ağız ilişkisi olmadığından dolayı- dikkate alınmadığı için graf herhangi bir geometrik şekil ifade etmemektedir.



Şekil 4: Graf Örneği

Şekil 3'te işlemlenmiş olan Dijkstra algoritmasında merkez olarak bir Güney Azerbaycan Türkçesi değişkisi olan Couzeh ağzı temel alınmıştır. Bu temel çerçevesinde Dijkstra algoritması temelli işlem adımları şu şekilde oluşmuştur:

1. Adım: Dijkstra algoritmasının uygulanmasından Couzeh ağzının köşe değeri 0, Couzeh ağzından doğrudan ulaşılabilen ağızların değerleri LUA algoritmasına göre aralarındaki ölçülen değerlerdir. Couzeh ağzı ile doğrudan bağlantılı olmayan ağızlar ise ∞ değerini alır. Sonsuz ifadesi henüz yolun minimal uzaklığının saptanmadığı anlamına gelir. O halde aşağıdaki tablo verilebilir:

Tablo 5: 1. Adım Değerleri

	Khel- tabad	Kara- su	Beha- ristan	Mensu- rabad	Mehr-i Zemin	Muşe- kiye	Şaneg	Cou- zeh
Couzeh	∞	2,04	∞	2,35	∞	1,99	∞	0

2. Adım: Couzeh ağzı ile Muşekiye ağzı arasındaki uzaklığa bakıldığında Muşekiye ağzından doğrudan ulaşılabilen ağızlar Couzeh, Karasu ve Beharistan ağızlarıdır (bkz. Şekil 4). Muşekiye ve Couzeh ağızları arası fark 1,99 dur. Fakat daha önce Couzeh ağzı 0 değeri aldığından ve yeni değer $0 < 1,99$ olduğu için Couzeh ağzının köşe değerini artıracığından Couzeh ağzı için köşe değeri değiştirilmez.

Bu bağlamda, Muşekiye ağzından Karasu ağzına gidilmesi için farklar toplamı Couzeh-Muşekiye-Karasu arası farklar toplamı olup bu değer $1,99+0,86=2,85 > 2,04$ olarak karşımıza çıkar ve yeni değer Karasu köşesinin değerini artırır dolayısıyla bu ağızda işlem yapılmaz.

Muşekiye ağzından doğrudan ulaşılabilen Beharistan ağzına uzaklıklar toplamı $1,99+0,95 = 2,94 < \infty$ olduğundan Beharistan köşesinin yeni değeri 2,94 olarak karşımıza çıkar. Bu işlemler sonucunda adım değerleri Tablo 6'da gösterildiği biçimde güncellenebilir:

Tablo 6: 2. Adım Değerleri

	Khelta- bad	Kara- su	Beha- ristan	Mensu- rabad	Mehr-i Zemin	Muşe- kiye	Şaneg	Couzeh
Couzeh	∞	2,04	2,94	2,35	∞	1,99	∞	0

3. Adım: Muşekiye ağzı ile Beharistan ağzı arasındaki uzaklık incelendiğinde Beharistan ağzından doğrudan ulaşılabilen ağızlar Şanegh ve Kheltabad ağızlarıdır. Şanegh ağzı için kenarlar Couzeh-Muşekiye-Beharistan-Şanegh olup bu ağızlar arasındaki farklar toplamı $1,99+0,95+1,52 = 4,46 < \infty$ olup Şanegh ağzının yeni değeri 4,46 olarak karşımıza çıkar.

Kheltabad ağızı için ise kenarlar Couzeh-Muşekiye-Beharistan-Kheltabad olup bu ağızlar arasındaki farklar toplamı $1,99+0,95+1,14 = 4,08 < \infty$ olup Kheltabad ağızının yeni değeri 4,08 olur. Böylece bu hesaplar sonrasında ortaya çıkan adım değerleri tablosu aşağıdaki gibi güncellenebilir:

Tablo 7: 3. Adım Değerleri

	Khelta- bad	Kara- su	Beharis- tan	Mensu- rabad	Mehr-i Zemin	Muşe- kiye	Şaneg	Couzeh
Couzeh	4,08	2,04	2,94	2,35	∞	1,99	4,46	0

4. Adım: Beharistan ağızı ile Kheltabad ağızı arasındaki uzaklık incelendiğinde Kheltabad ağızından doğrudan ulaşılabilen ağız Mehr-i Zemin'dir. Mehr-i Zemin için kenarlar Couzeh-Muşekiye-Beharistan-Kheltabad-Mehr-i Zemin olup bu ağızlar arasındaki farkın toplamı $1,99+0,95+1,14+1,28=5,36 < \infty$ 'dur. Buradan Mehr-i Zemin için yeni değer 5,36 olur ve bu ağza ilişkin bilgiler Tablo 8'deki gibi güncellenir:

Tablo 8: 4. Adım Değerleri

	Khelta- bad	Kara- su	Beharis- tan	Mensu- rabad	Mehr-i Zemin	Muşe- kiye	Şaneg	Couzeh
Couzeh	4,08	2,04	2,94	2,35	5,36	1,99	4,46	0

5. Adım: Mehr-i Zemin ağızından ulaşılabilen tek ağız Mensurabad'dır. Mehr-i Zemin'in değerinin 5,36 olduğu dikkate alınırsa Mensurabad köyünün değeri $5,36+1,24=6,6$ olur. Mensurabad'ın değeri daha önce 2,35 olarak bulunmuştu. Bu değer ilk değerden büyük olduğundan işlem yapılmaz.

Tablo 9: 5. Adım Değerleri

	Khel- tabad	Kara- su	Beha- ristan	Mensu- rabad	Mehr-i Zemin	Muşe- kiye	Şaneg	Couzeh
Couzeh	4,08	2,04	2,94	2,35	5,36	1,99	4,46	0

6. Adım: Mensurabad ağızı açısından durum incelendiğinde bu ağızdan doğrudan ulaşılabilen ağızlar Couzeh ve Karasu ağızlarıdır. Mensurabad ile Couzeh ve Karasu ağızları arasındaki uzaklık farkının toplamı ise Couzeh ve Karasu ağızlarının değerlerini artırır. Dolayısıyla işlem yapılması mümkün olmaz.

Tablo 10: 6. Adım Değerleri

	Khel- tabad	Kara- su	Beha- ristan	Mensu- rabad	Mehr-i Zemin	Muşe- kiye	Şaneg	Couzeh
Couzeh	4,08	2,04	2,94	2,35	5,36	1,99	4,46	0

7. Adım: Bu işlemler içerisinde yukarıda 2. adımda Couzeh ağzı ile Muşekiye ağzı arasındaki uzaklık hesaplanmıştı. Bu adımda da Couzeh ağzı ile Karasu ağzı arasındaki uzaklığın incelendiği durumda Karasu ağzından doğrudan ulaşılabilen Couzeh, Muşekiye ve Mensurabad ağızlarına değerler daha önce ulaşılan değerlerden büyük olacağından bu noktada işlem yapılmaz.

Tablo 11: 7. Adım Değerleri

	Khel- tabad	Kara- su	Beha- ristan	Mensu- rabad	Mehr-i Zemin	Muşe- kiye	Şaneg	Couzeh
Couzeh	4,08	2,04	2,94	2,35	5,36	1,99	4,46	0

8. Adım: Burada da yine bu işlemler içerisinde yukarıda 2. adımda Couzeh ağzı ile Muşekiye ağzı arasındaki uzaklık hesaplanmıştı. Bu adımda da Couzeh ağzı ile Mensurabad ağzı arasındaki uzaklığın hesabında Mensurabad ağzından Mehr-i Zemin ve Karasu ağızlarına gidilebilir.

Karasu için elde edilen yeni değer ilk değerden büyük olacağından işlem yapılması mümkün görünmemektedir.

Mehr-i Zemin ağzı için ise Couzeh-Mensurabad-Mehr-i Zemin ağızları arasındaki farkın toplamı $2,35+1,24=3,59$ olup ilk değerden küçüktür. Bu Mehr-i Zemin için yeni değer anlamına gelmektedir.

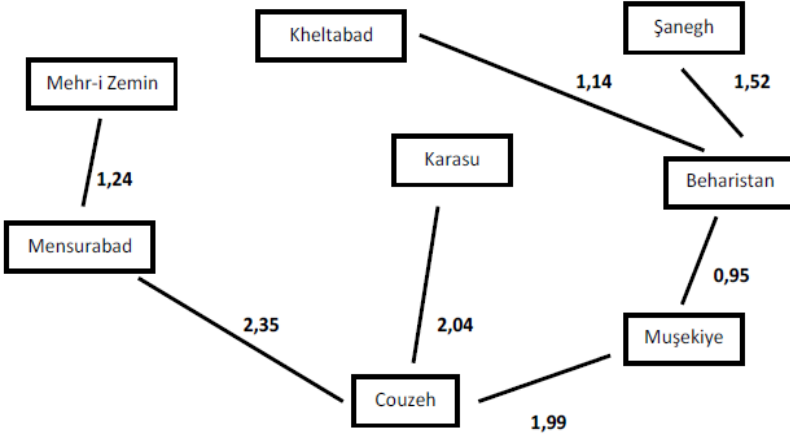
Dolayısıyla tablo aşağıdaki şekilde güncellenmiştir:

Tablo 12: 8. Adım Değerleri

	Khelta- bad	Kara- su	Beha- ristan	Mensu- rabad	Mehr-i Zemin	Muşe- kiye	Şaneg	Couzeh
Couzeh	4,08	2,04	2,94	2,35	3,59	1,99	4,46	0

9. Adım: 3. Adımda Muşekiye ağzından Beharistan ağzına minimal uzaklık ölçülmüştür. Bu adımda Muşekiye ağzı ile Karasu ağzı arasındaki uzaklık ölçüldüğünde ve Muşekiye'den Karasu ağzına ulaşmaya çalışılması durumunda 7. adımda olduğu gibi Karasu ağzından doğrudan ulaşılabilen Couzeh, Muşekiye ve Mensurabad ağızlarındaki değerler daha önce ulaşılan değerlerden büyük olacağından bu noktada da işlem yapılmaz.

Tüm bu veriler ışığında Dijkstra algoritması çözümlemesi sonucunda ağızlar arasındaki ilişki tüm işlemlerin neticesinde elde edilen son adımda Şekil 4'te ulaşılabilen ağaç grafiğiyle somutlaştırılmıştır:



Şekil 5: Dijkstra Uzaklığını Gösteren Son Ağaç Örneği

Şekil 4’te ortaya çıkan ağaç, Şekil 3’te sunulmuş olan grafin Dijkstra algoritmasına uygulanmış adımlarının neticesidir.

Tartışma ve Sonuç

Bu çalışma, İran’ın Merkezi ve Kum vilayetlerinde konuşulmakta olan iki Türk dili değişkesinin -Halaç Türkçesi ve Güney Azerbaycan Türkçesi- sözlüksel uzaklığını Leipzig-Jakarta çekirdek sözcük listesindeki (Tadmor ve Haspelmath, 2009) maddeler temelinde ağız ölçüsel (diyalektometrik) ve algoritmik yöntemler ışığında saptamayı amaçlamıştır.

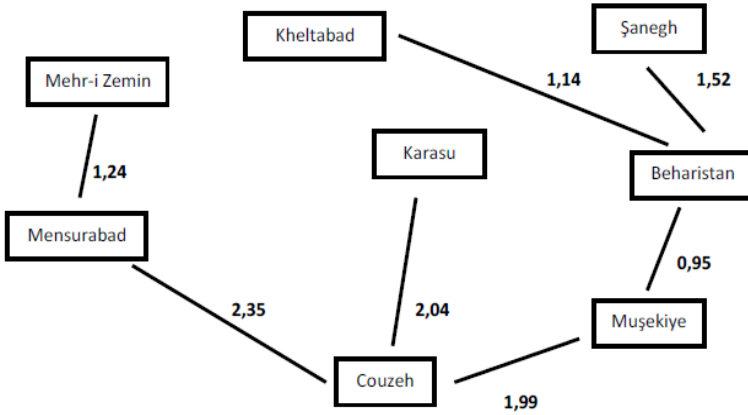
Bu bağlamda Levenshtein uzaklık algoritması temelinde GAT Couzeh ağızının Halaç Türkçesi ağızları arasındaki uzaklığı şu sırayla ortaya çıkmıştır:

- Muşekiye (kuzey doğu) ağızı: $d(\text{Couzeh}, \text{Muşekiye}) = 1,99$;
- Karasu (ana ağız) ağızı: $d(\text{Couzeh}, \text{Karasu}) = 2,04$;
- Beharistan (merkezi ağız) ağızı: $d(\text{Couzeh}, \text{Beharistan}) = 2,14$;
- Mensurabad (merkez doğu) ağızı: $d(\text{Couzeh}, \text{Mensurabad}) = 2,35$;
- Şanegh (güney) ağızı: $d(\text{Couzeh}, \text{Şanegh}) = 2,38$;
- Mehr-i Zemin (kuzey) ağızı: $d(\text{Couzeh}, \text{Mehr-i Zemin}) = 2,48$;
- Kheltabad (batı) ağızı: $d(\text{Couzeh}, \text{Kheltabad}) = 2,54$.

Bu çözümlemede Couzeh ağızı ile Halaç Türkçesinin Muşekiye (kuzey doğu) ağızı (=1,99) arasındaki sözlüksel uzaklık en az iken Kheltabad ağızı ile uzaklığın en çok (=2,54) olduğu belirlenmiştir. Batı ağızının konu-

şulduğu bölge her ne kadar İran'da Azerbaycan Türkçesinin konuşulduğu bölgeye yakın olsa da Couzeh ağız ile uzaklığı daha çok hesaplanmıştır. Bu durum Couzeh ağızının Halaç Türkçesi konuşulan bölgenin ortasında kalmasından kaynaklanmış olabilir. Couzeh'nin bu konumda olması Halaç Türkçesinin kenarda kalan ağızları ile uzaklığının daha fazla olmasında da etkili olmuş olabilir

Halaç Türkçesi ağızlarının Couzeh ağız ile ilişkisini tespit etmek üzere söz konusu Levenshtein algoritma verileri sonrasında Dijkstra algoritmasıyla da incelenmiş ve ilişkili graf örneği (bkz. Şekil 3) ağaç (Şekil 4) çıkartılmıştır. Burada vurgulanması gereken nokta Levenshtein algoritması iki ağız arasındaki uzaklığı ortaya çıkartırken Dijkstra algoritması ise GAT değişkesi Couzeh merkezinde yedi Halaç Türkçesi ağız arasındaki ilişkiyi incelemektedir. Dolayısıyla bu çalışma, üç ve üzeri değişkenin arasındaki ilişkinin incelenebilmesi için Dijkstra algoritmasının ağız çalışmalarına uyarlanması bir örneğini teşkil etmiştir. Bu yönüyle söz konusu bu çalışma çağdaş Türk lehçelerinin ağızları arasındaki uzaklığı ve algoritmik ilişkilerini disiplinlerarası bir yaklaşımla çözümlemeyi amaçlaması bağlamında yenilikçi bir yaklaşım benimsemiştir. Tüm bu veriler ışığında Dijkstra algoritması çözümlemesi sonucunda ağızlar arasındaki ilişki tüm işlemlerin neticesinde elde edilen son adımda ulaşılabilen ağaç grafiğiyle somutlaştırılmıştır:



Şekil 6: Dijkstra Algoritması İle Elde Edilen Ağaç Graf Örneği

Bu yönüyle son dönemde algoritma temelli incelemelerin ağız ölçümsel (diyalektometrik) incelemeleri, hızla gelişmekte olan yapay zekâ (İng. artificial intelligence) çalışmalarıyla desteklenmeye açık hâle gelecektir. Bu çalışma; Türk dili, matematik ve yapay zekâ disiplinlerini dil inceleme alanına tanıtmaya potansiyeline sahip olması açısından özgün bir yere sahiptir.

Kaynakça

- Anonby, E. (2012). Lori language ii. Sociolinguistic Status of Lori. *Encyclopædia Iranica*. <https://iranicaonline.org/articles/lori-language-ii>
- Bolognesi, R. ve Heeringa, W. (2002). De invloed van dominante talen op het lexicon en de fonologie van Sardische dialecten. *Gramma/TTT, 1: tijdschrift voor taalwetenschap*, 9(1), 45-84.
- Çam, A. (2021). *Şahseven Türkçesi* [Yayımlanmamış doktora tezi]. İstanbul Kültür Üniversitesi.
- Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik, 1*, 269-271.
- Doerfer, G. Hesch, W., Scheinhardt, H. ve Tezcan, S. (1971). *Khalaj materials*. Indiana University Press.
- Doerfer, G. (1987). *Lexik und Sprachgeographie des Chaladsch. 2 Volumes*. Harrassowitz.
- Doerfer, G. (1988). *Grammatik des Chaladsch* (Turcologica 4). Harrassowitz.
- Doerfer, G. (1998). Turkic languages of Iran. L. Johanson & É. Á. Csató (Eds.) içinde, *The Turkic Languages* (s. 273-282). Rotledge.
- Dreyfus, S. E. (1969). An appraisal of some shortest-path algorithms. *Operations research*, 17, 395-412.
- Eberhard, D. M., Gary F. S. & Charles D. F. (Eds.). (2022). *Ethnologue: Languages of the world*. Twenty-fifth edition. SIL International. Online version: <http://www.ethnologue.com>.
- Goebel, H. (1982). *Dialektometrie; Prinzipien und Methoden des Einsatzes der Numerischen Taxonomie im Bereich der Dialektgeographie*. Akademie der Wissenschaften.
- Goebel, H. (2017). Dialectometry. Ch. Boberg, J. A. Nerbonne & D. Watt (Eds.), *The Handbook of Dialectology* içinde (s. 123-142). Wiley-Blackwell.
- Gooskens, C. & Heeringa, W. (2004). Perceptive evaluation of Levenshtein dialect distance measurements using Norwegian dialect data. *Language Variation and Change*, 16(3), 189-207.
- Gudschinsky, S. C. (1956). The ABC's of lexicostatistics (glottochronology). *Word*, 12(2), 175-210.
- Haspelmath, M. & Tadmor, U. (Eds.). (2009). *Loanwords in the World's Languages: A Comparative Handbook*. Walter de Gruyter. <https://doi.org/10.1515/9783110218442.55>
- Heeringa, W. J. (2004). Measuring dialect pronunciation differences using Levenshtein distance. *Unpublished Doctoral Dissertation*. University Library Groningen.
- Heeringa, W. & Prokić, J. (2017). Computational dialectology. C. Boberg, J. A. Nerbonne, & D. Watt (Eds.), *The Handbook of Dialectology* içinde (s. 330-347). Wiley-Blackwell.

- Kuribayashi, Y. (2012). Grammaticalized topics in Kashkay: The implication for the relativization of Turkic languages. *Studia Uralo-Altica*, 49, 311-318.
- Minorsky, M. (1940). The Turkish Dialect of the Khalaj, *Bulletin of the Society of Oriental and African Studies* 10, 417-37.
- Moğaddäm, M. (1318). *Guyeşha-ye Wäfs wä Täfräş*, İrañ-Kudé 11 (Teheran, h.ş.).
- Nerbonne, J. & Heeringa, W. (2010). Measuring dialect differences. P. Auer & J. E. Schmidt (Eds), *An International Handbook of Linguistic Variation. Theories and Methods Volume 1* içinde (s. 550-566). De Gruyter Mouton. <https://doi.org/10.1515/9783110220278>
- Nerbonne, J. & Siedle, C. (2005). Dialektklassifikation auf der Grundlage aggregierter Ausspracheunterschiede. *Zeitschrift für Dialektologie und Linguistik*, 129-147.
- Osenova, P., Heeringa, W. & Nerbonne, J. (2009). A quantitative analysis of Bulgarian dialect pronunciation. *Zeitschrift für slavische Philologie*, 66(2), 425-458.
- Séguy, J. (1973). La dialectométrie dans l'Atlas linguistique de la Gascogne. *Revue de linguistique romane*, 37, 1-24.
- Ruohonen, K. (2008). *Graph theory*. Tampere University of Technology.
- Szmrecsanyi, B. (2013). *Grammatical variation in British English dialects: A study in corpus-based dialectometry*. Series: Studies in English Language. Cambridge University Press.
- Tadmor, U. (2009). Loanwords in the world's languages: Findings and results. M. Haspelmath & U. Tadmor (Eds.), *Loanwords in the World's Languages: A Comparative Handbook* içinde (s. 55-75). Walter de Gruyter. <https://doi.org/10.1515/9783110218442.55>.
- Tekin, T. (1989). Türk dili diyalektlerinin yeni bir tasnifi. *Erdem*, 5(13), 141-168.
- Whiting, P. D. & Hillier, J. A. (1960). A method for finding the shortest route through a road network, *Operations Research Quarterly*, 11, 37-40.
- Wieling, M., Prokić, J. & Nerbonne, J. (2009). Evaluating the pairwise string alignment of pronunciations. *Proceedings of the EACL 2009 Workshop on Language Technology and Resources for Cultural Heritage, Social Sciences, Humanities, and Education (LaTeCH-SHELT&R 2009)* içinde (s. 26-34).
- Windfuhr, G. (Ed.). (2009). *The Iranian languages*. Routledge.
- Windfuhr, G. & Perry, J. R. (2009). Persian and Tajik. G. Windfuhr (Ed.) içinde, *The Iranian Languages* (s. 416-545). Routledge.

Extended Summary

Measuring the phonetic and lexical distances of Turkic languages are mostly neglected in the relevant literature (see Akkuş, 2021). Drawing on a quantitative approach, this study thus aims at discussing the use of the Levenshtein Distance Algorithm (LDA) and Dijkstra Algorithm (DA) in measuring (i) phonetic and (ii) lexical distances of a number of Turkic languages, namely Khalaj Turkic and South Azerbaijani to discuss the availability of a relatively valid means to measure the rates of intelligibility for Turkic varieties, based on the Leipzig-Jakarta core vocabulary list (Tadmor and Haspelmath, 2009). To this end, all Khalaj dialects spoken in Markazī and Qom provinces, as classified by Doerfer (1988), and the Couzeh variety of South Azerbaijani were compared in the present study. LDA, a unigram-based string metric that compares two sequences and measures their differences, allows researchers to measure “individual speech segments” while Dijkstra Algorithm consider the relationship of three or more variables. The database was formed based on the 100-item Leipzig-Jakarta list (Haspelmath and Tadmor, 2009). The data was obtained through direct elicitations from native speaker language consultants of the Turkic languages under scrutiny. An algorithmic analysis was applied to identify the most valid method to measure the relative distances.

The study also investigates whether there is any impact of geographical proximity (areal convergence) between Khalaj (Doerfer, 1971, 1987, 1988; Tekin, 1989) and South Azerbaijani. The data used in the study were compiled from all Khalaj-speaking settlements in Markazi and Qom provinces in Iran, and from the village of Couzeh, where South Azerbaijani Turkic is spoken in the province of Qom, during the fieldwork carried out between July and September in 2021. Within the scope of this study, the lexical distance of two Turkic varieties spoken in the same language ecology was measured in an algorithmic framework, and the values of the phonological distances of the lexical items were measured.

Based on the findings of the Levenshtein distance algorithm, the distance between South Azerbaijani dialect and Khalaj Turkic dialects emerged in the following order:

Muşekiye (north east) dialect:	$d(\text{Couzeh}, \text{Muşekiye}) = 1,99;$
Karasu (main) dialect:	$d(\text{Couzeh}, \text{Karasu}) = 2,04;$
Beharistan (central) dialect:	$d(\text{Couzeh}, \text{Beharistan}) = 2,14;$
Mensurabad (central east) dialect:	$d(\text{Couzeh}, \text{Mensurabad}) = 2,35;$
Şanegh (south) ağızı dialect:	$d(\text{Couzeh}, \text{Şanegh}) = 2,38;$
Mehr-i Zemin (North) dialect:	$d(\text{Couzeh}, \text{Mehr-i Zemin}) = 2,48;$
Kheltabad (west) dialect:	$d(\text{Couzeh}, \text{Kheltabad}) = 2,54.$

In the light of the data in the table, it is seen that the Khalaj dialect with the lowest lexical distance from the Couzeh dialect according to the Levenshtein algorithm is the Musekiye (north east) dialect (LUA=1.99). Afterwards, the dialect most converging to the Couzeh dialect was determined as the Karasu

(main dialect) dialect with the result of $d(\text{Couzeh}, \text{Karasu}) = 2.04$. . The distance between Beharistan (central dialect) and Couzeh dialect was determined as 2.14. While the Mensurabad (central east) dialect (LUA=2.35) and Şaneğh (south) dialect (LUA=2.38) diverge with the Couzeh dialect in terms of lexical distance with close values, Mehr-i Zemin (north) (LUA=2, 48) and Kheltabad (west) (LUA=2,54) were calculated as the farthest dialects. Since the Levenshtein algorithm examines the difference and/or similarity between the two strings, in this analysis, the aforementioned data were also examined with the Dijkstra algorithm in order to determine the relationship of all Khalaj dialects with the Couzeh dialect with multidimensional analysis and graph samples were extracted. The point that should be emphasized here is that Levenshtein algorithm reveals the distance between two dialects, while Dijkstra algorithm examines the relationship between seven Khalaj dialects in the center of GAT variable Couzeh.

In this analysis, it was revealed that the lexical distance between the Couzeh dialect and the Muşekiye (north east) dialect of the Khalaj language (=1.99) was the least, while the distance from the Kheltabad dialect was the most (=2.54). Although the region where the western dialect is spoken is close to the region where Azerbaijani Turkic is widely spoken in Iran, its distance from the Couzeh dialect has been calculated more. In this case, it may have been due to the fact that the Couzeh dialect was in the middle of the Khalaj-speaking region. This location of Couzeh also seems to be effective in the fact that it is more distant from the marginal dialects of Khalaj.

In order to determine the relationship between all the Khalaj dialects and the Couzeh dialect, Levenshtein algorithm data was also correlated with the Dijkstra algorithm, and relevant graphs and trees were extracted, accordingly. It is worth stating that Levenshtein algorithm reveals the distance between two dialects, while Dijkstra algorithm examines the relationship between seven Khalaj dialects, placing the GAT variety in the center. In this respect, this study has adopted an innovative approach in the sense that it aims to analyze the distance and algorithmic relations between the dialects of contemporary Turkic languages through interdisciplinary lens. In the light of all these data, the relationship between the dialects as a result of the Dijkstra algorithm analysis is concretized with the tree graph that can be reached in the last step obtained as a result of all the operations.

EK 1: Leipzig-Jakarta Çekirdek Sözcük Listesi

	Türkçe	İngilizce	Dil bilgisi kategorisi	GAT (Couzeh değişkesi)	Halaç Türkçesi (Beharistan)
1	ateş	'fire'	ad	/ot/	/huot/
2	burun	'nose'	ad	/burun/	/burun/
3	gitmek	'to go'	eylem	/gedmek/	/varmaq/
4	su	'water'	ad	/su/	/suv/
5	ağız	'mouth'	ad	/ayz/	/ayuz/
6	dil	'tongue'	ad	/dil/	/til/
7	kan	'blood'	ad	/qon/	/qon/
8	kemik	'bone'	ad	/simik/	/simik/
9	sen	'2sg pronoun'	adıl	/sæn/	/sæn/
10	kök	'root'	ad	/riʃe/	/riʃe/
11	gelmek	'to come'	eylem	/gelmek/	/kelmek/
12	meme	'breast'	ad	---	---
13	yağmur	'rain'	ad	/jayıf/	/jayıf/
14	ben	'1sg pronoun'	adıl	/mæn/	/mæn/
15	ad	'name'	ad	/v:t/	/v:t/
16	bit	'louse'	ad	/bit/	/bit/
17	kanat	'wing'	ad	/pær/	/pær/
18	et	'flesh/meat'	ad	/æt/	/æt/
19	kol	'arm/hand'	ad	/bazu/	/qol/
20	sinek	'fly'	ad	---	/quduqu/
21	gece	'night'	ad	/gedze/	/kietfe/
22	kulak	'ear'	ad	/qulaq/	/qulbq/
23	ense	'neck'	ad	/boyn gerisi/	/bu:n artuu/
24	uzak	'far'	önad	/irax/	/huraq/
25	yapmak	'to do/make'	eylem	/endzam vermek/	/endzam viermek/
26	ev	'house'	ad	/ev/	/hæv/
27	taş/kaya	'stone/rock'	ad	/dɒʃ/	/tɒʃ/
28	acı	'bitter'	önad	/adzi/	/hatʃuy/
29	söylemek	'to say'	eylem	/demek/	/heymaq/
30	diş	'tooth'	ad	/diʃ/	/tiʃ/
31	saç	'hair'	ad	/dik/	/tik/

32	büyük	'big'	önad	/bedik/	/bydyk/
33	bir	'one'	önad	/bi:/	/bi:/
34	kim	'who'	adıl	/kim/	/kim/
35	o	'3sg pronoun'	adıl	/u:/	/u:/
36	vurmak	'to hit/beat'	eylem	/vurmaq/	/vurmaq/
37	bacak/ ayak	'leg/foot'	ad	/ajax/	/hadaq/
38	boynuz	'horn'	ad	/föxe/	/föxe/
39	bu	'this'	adıl	/bu/	/bu/
40	balık	'fish'	ad	/møhi/	/møhi/
41	dün	'yesterday'	belirteç	/dinen/	/engir/
42	içmek	'to drink'	eylem	/iřmek/	/itřmek/
43	siyah	'black'	önad	/qare/	/qara/
44	göbek deliđi	'navel'	ad	/nøf/	/kyndyk/
45	durmak	'to stand'	eylem	/durmax/	/turmaq/
46	ısırmak	'to bite'	eylem	/qøz vurmx/	/qøzlemøq/
47	arka	'back'	önad	/dal/	/hart/
48	rüzgar	'wind'	ad	/yel/	/yiel/
49	duman	'smoke'	ad	/titin/	/tytyn/
50	ne	'what?'	adıl	/neme/	/nese/
51	çocuk	'child'	ad	/uřøx/	/qpl/
52	yumurta	'egg'	ad	/numurte/	/jumurqa/
53	vermek	'to give'	eylem	/vermek/	/viermek/
54	yeni	'new'	önad	/jengi/	/jengi/
55	yanmak	'to burn (intr.)'	eylem	/janmax/	/kienmek/
56	deđil	'not'	edat	/yox/	/daq/
57	iyi	'good'	önad	/jaxi/	/havul/
58	bilmek	'to know'	eylem	/bi:lmek/	/bi:lmek/
59	diz	'knee'	ad	/diz/	/ti:z/
60	kum	'sand'	ad	/řun/	/řuna/
61	gölmek	'to laugh'	eylem	/gilmek/	/kylmek/
62	duymak	'to hear'	eylem	/eřitmek/	/eřitmek/
63	toprak	'soil'	ad	/turpax/	/turpøq/
64	yaprak	'leaf'	ad	/berg/	/belk/

Halaç Türkçesinin ağızları ile Güney Azerbaycan Türkçesinin Kum (Couzeh) ağızı arasındaki sözlüksel uzaklık: Levenshtein ve Dijkstra algoritmaları amalgamı

65	kırmızı	'red'	önad	/qirmiz/	/qizil/
66	karaciğer	'liver'	ad	/kæbæd/	/qaracigær/
67	saklamak	'to hide'	eylem	/qaym eylæmæk/	/kieʃitmek/
68	deri	'skin/hide'	ad	/dæri/	/teri/
69	emmek	'to suck'	eylem	/mukilæmæk/	/suormaq/
70	taşımak	'to carry'	eylem	/εji varmaq/	/jietmek/
71	karınca	'ant'	ad	/gumurgõtʃ/	/gumurgõtʃ/
72	ağır	'heavy'	önad	/aʃir/	/æʃir/
73	almak	'to take'	eylem	/almaq/	/almaq/
74	yaşlı	'old'	önad	/qodʒa/	/qarri/
75	yemek	'to eat'	eylem	/jεmek/	/jiemek/
76	but	'thigh'	ad	/but/	/but/
77	kalın	'thick'	önad	/yoʒur/	/yoʒun/
78	uzun	'long'	önad	/fasiledar/	---
79	üfleme	'to blow'	eylem	/puflæmek/	/puflamaq/
80	odun	'wood'	ad	/aʃatʃ/	/hagatʃ/
81	koşmak	'to run'	eylem	/qatʃmaq/	/tietzmeq/
82	düşmek	'to fall'	eylem	/diʃmek/	/tiʃmek/
83	göz	'eye'	ad	/giez/	/kez/
84	kül	'ash'	ad	/kil/	/kil/
85	kuyruk	'tail'	ad	/dom/	/qurduq/
86	köpek	'dog'	ad	/it/	/it/
87	ağlamak	'to cry/ weep'	eylem	/aʃlamaq/	/huʃlamaq/
88	bağlamak	'to tie'	eylem	/ba:ʃlamaq/	/va:maq/
89	görmek	'to see'	eylem	/giermek/	/kermek/
90	tatlı	'sweet'	önad	/ʃi:rin/	/ʃi:rin/
91	ip	'rope'	ad	/tænab/	/tænaf/
92	gölge	'shade/sha- dow'	ad	/saje/	/saje/
93	kuş	'bird'	ad	/pærændæ/	/pærændæ/

94	tuz	'salt'	ad	/tuz/	/tuz/
95	küçük	'small'	önad	/kuçuk/	/bætʃæ/
96	geniş	'wide'	önad	/bejik/	/kienk/
97	yıldız	'star'	ad	/ulduz/	/julduz/
98	içinde	'in'	belirteç	/itfidæ/	/itʃitʃæ/
99	sert	'hard'	önad	/aʃir /	/sæxt/
100	ezmek	'to crush/ grind'	eylem	/xurd elemek/	/xurdlamaq/

