

BLEU-dan COMET-ə: ərəb–türk audiovizual tərcüməsində ənənəvi və müasir qiymətləndirmə metriklərinin müqayisəli təhlili

Yasemin Yazal 

Xülasə. Tərcümə müasir dövrün əsas ehtiyaclarından biri kimi insanlar və mədəniyyətlər arasında mühüm ünsiyyət körpüsü rolunu oynayır. Tərcüməyə artan tələbat avtomatik tərcümə sistemlərinin inkişafını qaçılmaz etmiş və rəqəmsal dövrdə bu sistemlərin istifadəsi sürətlə genişlənməmişdir. Bu sistemlər tərcümə prosesini daha sürətli həyata keçirməyə imkan versə də, onların çıxışları mədəni transfer, semantik ardıcılıq, kontekstual adekvatlıq və ümumi mətn bütövlüyü baxımından hələ də müəyyən məhdudiyyətlər göstərə bilər. Bu səbəbdən maşın tərəfindən yaradılan tərcümələrin keyfiyyətini qiymətləndirmək üçün avtomatik qiymətləndirmə metrikləri getdikcə daha mühüm əhəmiyyət kəsb edir. Bu tədqiqat maşın tərcüməsinin qiymətləndirilməsində istifadə olunan ənənəvi metriklərin, yəni BLEU, ROUGE və METEOR-un, eləcə də müasir metriklərin, yəni BERTScore, LaBSE və COMET-in nəzəri və praktik baxımdan müqayisəli təhlilini təqdim edir. Ənənəvi metriklər əsasən namizəd və istinad mətnləri arasındakı səthi leksik uyğunluğa əsaslandığı üçün türk və ərəb kimi morfoloji baxımdan mürəkkəb dil cütliklərində semantik və kontekstual ekvivalentliyin qiymətləndirilməsində məhdud qala bilər. Buna görə də tədqiqatın empirik hissəsi müasir qiymətləndirmə metrikləri vasitəsilə aparılan çoxdillə təhlilə yönəlmişdir. Nəticələr göstərir ki, Language Reactor altyazı tərcümələri mənbə mətnlə semantik uyğunluq baxımından müasir metriklərdə çox vaxt daha yüksək nəticələr əldə edir, insan tərcümələri isə mədəni transfer, axıcılıq və kontekstual təbiilik baxımından daha uğurlu nəticələr ortaya qoyur.

Açar sözlər: maşın tərcüməsi; COMET; BERTScore; LaBSE; language reactor

Kırıkkale University, doktorant, Kırıkkala, Türkiyə

E-poçt: yaseminyazal9@gmail.com

Daxil oldu: 29 Yanvar 2026; Qəbul edildi 21 May 2026; Onlayn dərc edildi: 28 İyun 2026

© Müəllif(lər) 2026. Bu, Creative Commons Attribution-NonCommercial 4.0 Beynəlxalq Lisenziyası (CC BY-NC 4.0) şərtləri altında paylanan açıq girişli məqalədir.

From BLEU to COMET: A Comparative Analysis of Traditional and Modern Evaluation Metrics in Arabic–Turkish Audiovisual Translation

Yasemin Yazal 

Abstract. Translation, one of the fundamental needs of the modern era, serves as a vital bridge of communication between people and cultures. The growing demand for translation has made the development of automated translation systems inevitable, and their use has rapidly expanded in the digital age. Although these systems enable faster translation processes, their outputs may still remain limited in terms of cultural transfer, semantic consistency, contextual adequacy, and overall coherence. Consequently, automatic evaluation metrics have become increasingly important for assessing the quality of machine-generated translations.

This study provides a comparative theoretical and practical analysis of traditional metrics, namely BLEU, ROUGE, and METEOR, and modern metrics, namely BERTScore, LaBSE, and COMET, used in machine translation evaluation. Since traditional metrics mainly rely on surface-level lexical overlap between candidate and reference texts, they may remain limited in evaluating semantic and contextual equivalence in morphologically complex language pairs such as Turkish and Arabic. Therefore, the empirical section of this study focuses on multilingual analysis through modern evaluation metrics. The findings indicate that Language Reactor subtitle translations often achieve higher scores in modern metrics in terms of semantic alignment with the source text, whereas human translations yield more successful results in terms of cultural transfer, fluency, and contextual naturalness.

Keywords: machine translation; COMET; BERTScore; LaBSE; language reactor

Kırıkkale Universiteti, PhD Student, Kırıkkale, Turkey
E-mail: yaseminyazal9@gmail.com

Received: 29 January 2026; Accepted 22 May 2026; Published online: 28 June 2026

© The Author(s) 2026. This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Giriş

Tərcümə müxtəlif mədəniyyətlər arasında yeni bilik mənbələrinin və elmi sahələrin ötürülməsində mühüm rol oynayan əsas vasitələrdən biridir. Tarix boyunca bir çox sivilizasiya tərcümə vasitəsilə fərqli mədəni və elmi qaynaqlara çıxış əldə etmiş, bunun nəticəsində öz maariflənmə dövrlərini formalaşdırmışdır. Məlumdur ki, incəsənət, elm və fəlsəfə sahələri tərcümə sayəsində zənginləşmiş və daha geniş auditoriyalara çatmışdır (Göktürk, 1994). Yazılı dilin meydana çıxmasından etibarən insanlar ünsiyyəti təkmilləşdirmək üçün davamlı şəkildə müxtəlif yollar axtarmışdır. Heroqliflərlə başlayan yazılı ünsiyyət bu günün rəqəmsal audiovizual vasitələrinə qədər əhəmiyyətli dəyişikliklər keçirmişdir. Müasir dövrdə cəmiyyətlərin audiovizual tərcümə mərhələsinə tam şəkildə inteqrasiya olunduğu müşahidə edilir (Diaz Cintas, 2012). Rəqəmsal transformasiyaların sürətli inkişafı və insanlar arasında audiovizual ünsiyyətin geniş yayılması Tərcüməşünaslığın alt sahələrindən biri hesab edilən audiovizual tərcümənin inkişafını sürətləndirmişdir (Diaz Cintas & Remael, 2021). Bu gün ünsiyyətin getdikcə daha sosial xarakter alması ilə audiovizual tərcümə altyazı, dublyaj, üst yazı və audio təsvir kimi müxtəlif formaları əhatə edən daha geniş bir tədqiqat sahəsinə çevrilmişdir (Maestri, 2021). Qloballaşma şəraitində beynəlxalq şirkətlərin, xüsusilə biznes və xarici ticarət sahələrində qarşılaşdıqları əsas problemlərdən biri dil baryeridir. Bu problemin aradan qaldırılması istiqamətində uzun illərdir sistemli şəkildə müxtəlif təşəbbüslər irəli sürülür. Bu baryeri aradan qaldırmaq üçün inkişaf etdirilən aparıcı yanaşmalardan biri isə maşın tərcüməsidir (MT) (Yılmaz, 2023).

Maşın tərcüməsi

Maşın tərcüməsi (MT) həm nəzəri, həm də metodoloji mənada istifadə olunan bir anlayışdır. İlk olaraq, o, mətn və ya nitq girişinin mənbə dildən hədəf dilə ötürülməsində proqram əsaslı vasitələrin istifadəsini araşdıran kompüter dilçiliyinin bir alt sahəsi kimi müəyyən edilir. İkinci olaraq isə maşın tərcüməsi insan müdaxiləsi olmadan, proqram və ya tətbiqlər vasitəsilə məzmunun mənbə dildən hədəf dilə avtomatik şəkildə çevrilməsi prosesini ifadə edir. Bu kontekstdə insanın rolu yalnız sistem proqram təminatının hazırlanması və inkişaf etdirilməsi ilə məhdudlaşır (Sadıkov & Sarıgül, 2021).

Maşın tərcüməsi anlayışı ilk dəfə 1947-ci ildə, ilk kompüterin inkişafından bir il sonra Warren Weaver tərəfindən irəli sürülmüşdür. Ortaya çıxmasından bu günə qədər maşın tərcüməsi təbii dil emalı (NLP) sahəsinin ən mürəkkəb tədqiqat istiqamətlərindən biri kimi qiymətləndirilmişdir (Wang et al., 2022).

Maşın tərcüməsi maşın əsaslı sistemlərin bir dili başqa bir dilə çevirmək üçün necə istifadə edilə biləcəyini araşdıran tədqiqat sahəsidir. Warren Weaver tərəfindən 1947-ci ildə irəli sürülən MT anlayışı elektron kompüterlərin inkişafından bir il sonra meydana çıxmışdır (Wang et al., 2022).

Maşın tərcüməsi anlayışının iki əsas funksiyası vardır: Birincisi, o, mətn və ya nitqin dillərarası tərcümə prosesini proqram vasitələri baxımından araşdıran kompüter dilçiliyinin bir sahəsidir. İkincisi, proqram və ya tətbiqlər vasitəsilə hər hansı bir mətnin mənbə dildən hədəf dilə insan müdaxiləsi olmadan avtomatik şəkildə çevrilməsi prosesini ifadə edir (Sadıkov & Sarıgül, 2021). Elmi mühitdə ilk dövrlərindən etibarən maşın tərcüməsi əhəmiyyətli inkişaf yolu keçmişdir. Dünyanı dəyişdirən texnoloji yenilik dalğaları tərcümə sahəsini də yenidən formalaşdırmışdır (Barut, 2022).

Əlaqəli tədqiqatlar

Maşın tərcüməsinin qiymətləndirilməsi üçün inkişaf etdirilən ilk modellər əsasən n-qram uyğunluğu və səthi səviyyəli ekvivalentliyə əsaslanmışdır. Bu metriklər ənənəvi qiymətləndirmə metrikləri hesab edilir. İlk mərhələlərdə BLEU (Papineni et al., 2002), ROUGE (Lin, 2004) və METEOR (Lavie & Agarwal, 2007) kimi metriklər formal ekvivalentlik ilə insan tərcüməsi arasındakı korrelyasiyanı ölçmək üçün geniş şəkildə istifadə olunmuşdur. Lakin bu metriklərin eyni semantik məkan daxilində fəaliyyət göstərməməsi, semantik ekvivalentliyi və kontekstual mənanı tam şəkildə əhatə edə bilməməsi onların tənqid edilməsinə və sonradan təkmilləşdirilməsinə səbəb olmuşdur.

Bu məhdudiyyətlər tədqiqatçıları insan mühakiməsi ilə daha yüksək korrelyasiya göstərən müasir metriklər hazırlamağa yönəlmişdir. BERTScore (Zhang et al., 2020), LaBSE (Feng et al., 2022) və COMET (Rei et al., 2020) kimi metriklər yalnız səthi oxşarlığa əsaslanmaq əvəzinə, kontekstual və semantik uyğunluğu təmin edən, insan tərəfindən annotasiya edilmiş məlumatlar əsasında hazırlanmış modellər kimi ön plana çıxır. Son illərdə neyron əsaslı quruluşlarına görə bu müasir metriklər tərcümə keyfiyyətinin daha dəqiq qiymətləndirilməsi üçün ənənəvi metriklərdən daha çox üstün tutulur.

Əvvəlki tədqiqatlar nəzərdən keçirildikdə, müqayisəli təhlillərin əksəriyyətinin ingilis dili əsaslı dil cütlükləri üzərində aparıldığı, ərəb və türk kimi morfoloji baxımdan mürəkkəb dillərlə bağlı araşdırmaların isə məhdud qaldığı müşahidə edilir. Maşın tərcüməsi sistemləri əhəmiyyətli dərəcədə inkişaf etsə də, ərəb dili morfoloji zənginliyi, sintaktik çevikliyi, dialekt müxtəlifliyi və kontekstdən asılı məna quruluşu səbəbindən avtomatik tərcümə üçün çətin hədəf dil olaraq qalır. Bu dil xüsusiyyətləri, keyfiyyətli təlim məlumatlarının məhdudluğu ilə birlikdə, ərəb dilinə maşın tərcüməsi çıxışlarının dəqiqliyini və etibarlılığını azalda bilər (Slimani, 2025). Buna görə də bu tədqiqat Language Reactor əlavəsindən əldə edilən avtomatik altyazı tərcümələrini insan tərcümələri ilə müasir qiymətləndirmə metrikləri vasitəsilə müqayisə edərək bu boşluğu doldurmağı hədəfləyir.

Maşın tərcüməsi modelləri

Maşın tərcüməsi nəticələrinin keyfiyyəti istifadə olunan tərcümə modellərindən asılı olaraq dəyişir və modellər fərqləndikcə onların tərcümə performansını da dəyişiklik göstərir (Yılmaz, 2023).

Maşın tərcüməsi modelləri üç əsas kateqoriya üzrə təsnif edilir:

- *Qayda əsaslı maşın tərcüməsi (RBMT):*

Avtomatik tərcümədə tarixən istifadə olunan ilk metod kimi qayda əsaslı model mənbə dil məlumatlarını əvvəlcədən kodlaşdırılmış kompüter dilinə çevirir və onları hədəf dil kodlarına ötürür (Ayık Akça, 2022). Sintaktik, qrammatik və struktur uyğunluğu təmin etməklə bu model qarşılaşdırılan dillərin sistemlərini və cümlə quruluşlarını modelləşdirir, nəticədə sözbəsöz tərcümələr yaradır (Ramizoğlu, 2024). Lakin lüğətlərdə yer almayan sözlər və ya idiomatik ifadələrlə

qarşılaşdıqda onun performansını və tərcümə keyfiyyəti azalır, bu da səhvlərin yaranmasına səbəb olur (Yılmaz, 2023).

- *Statistik maşın tərcüməsi (SMT):*

Statistik maşın tərcüməsi ilk dəfə Brown et al. (1990) tərəfindən ehtimala əsaslanan bir yanaşma kimi irəli sürülmüşdür. Sistem geniş ikidilli məlumat bazalarından söz ekvivalentliyi ehtimallarını hesablayaraq rəqəmsal proqnozlaşdırma yolu ilə tərcümə həyata keçirir. Bu ehtimalları təhlil etməklə verilmiş cümlə üçün ən mümkün tərcüməni avtomatik şəkildə yaradır (Brown et al., 1990). Statistik model geniş ikidilli məlumat bazaları üzərində işləyir və mənbə cümləni hədəf dilə ən ehtimal olunan qarşılıq cümlə əsasında tərcümə edir (Banitz, 2020).

Statistik model geniş ikidilli məlumat bazaları üzərində işləyir və mənbə cümləni hədəf dilə ən ehtimal olunan qarşılıq cümlə əsasında çevirir. Statistik maşın tərcüməsi iki alt qrupa ayrılır: söz əsaslı model və ifadə əsaslı model (Banitz, 2020).

- *Neyron maşın tərcüməsi (NMT):*

Bu model günümüzdə ən inkişaf etmiş tərcümə sistemlərindən biri hesab edilir. Onun fərqləndirici xüsusiyyəti öyrənmə qabiliyyətinə malik olmasıdır (Yılmaz, 2023, s. 344). Neyron maşın tərcüməsi modeli əvvəlcə mənbə mətni bütövlükdə oxuyur və təhlil edir, daha sonra cümlənin kontekstinə uyğun olaraq hədəf cümləni söz-söz yenidən qurur (Wang et al., 2022).

Language Reactor

Texnologiyanın inkişafı və süni intellekt vasitələrinin geniş yayılması tərcümə və dil öyrənmə proseslərini asanlaşdıran platformaların inkişafına şərait yaratmışdır. Bu gün tərcümə sahəsində bacarıqlarını inkişaf etdirmək istəyən fərdlər və peşəkarlar üçün çoxsaylı rəqəmsal tərcümə vasitələri hazırlanmışdır. Bu sahədə inkişaf etdirilən və geniş istifadə olunan vasitələrdən biri Language Reactor (LR) əlavəsidir. İnnovativ veb-brauzer əlavəsi kimi Language Reactor audiovizual tərcümə məhsulları vasitəsilə istifadəçilərə həm avtomatik tərcümə, həm də dil öyrənmə imkanları təqdim edir, eyni zamanda tədqiqatçılara akademik araşdırmalarda məlumat toplamaq və müqayisəli təhlillər aparmaq imkanı yaradır.

Language Reactor Netflix, YouTube və digər onlayn yayım platformalarında fərqli dillərdə altı yazılarla məzmun izləməyə imkan verən brauzer əlavəsi kimi hazırlanmışdır. Bu əlavə dil öyrənmə prosesini asanlaşdırmaq məqsədilə bütün dil bacarıqlarını bir araya gətirir. İlk dövrlərdə Language Reactor with Netflix və Language Reactor with YouTube adları ilə təqdim edilən əlavə son dövrlərdə sadəcə Language Reactor adı ilə tanınır. O, Google Chrome brauzerində əlavə kimi quraşdırılır.

Bu əlavə vasitəsilə istifadəçilər audiovizual məzmunu dil səviyyələrinə uyğun şəkildə filtrləyə bilirlər. Bu mühtidə istifadəçilər sözlərin mənalərini ətraflı şəkildə görə və onların tələffüzünü dinləyə bilirlər. Language Reactor ani tərcümə ehtiyacını qarşılıyır. O, yalnız dil öyrənmək istəyən şəxslər üçün deyil, eyni zamanda bilmədikləri dildə audiovizual məhsulları izləmək istəyən istifadəçilər və eşitmə məhdudiyəti olan şəxslər üçün də faydalıdır; çünki bütün məzmun izləmə üçün altı yazılarla təqdim edilir.

Əlavə həmçinin kitablar, podkastlar və veb-saytlar üçün tərcümə və öyrənmə dəstəyi təqdim edir. İkidilli altı yazı və lüğət dəstəyi sayəsində Language Reactor tərcümə müqayisələrini xeyli asanlaşdırır və istifadəçilərə həm hədəf dili, həm də mənbə/ana dili eyni vaxtda görmək imkanı verir. LR fərdiləşdirilə bilən istifadəçi interfeysi, oxutmaya nəzarət, Chatbot Aria və PhrasePump kimi xüsusiyyətlər təqdim etməklə dil öyrənmə təcrübəsini daha cəlbedici və fərdiləşdirilmiş hala gətirir.

Qiymətləndirmə metrikləri

Müasir dövrdə tərcümə yalnız mənanın mənbə dildən hədəf dilə ötürülməsi ilə məhdudlaşmır; o, həm də avtomatik tərcümə vasitələrinin iş prinsiplərini və tərcümə keyfiyyətinin qiymətləndirilməsi prosesini əhatə edir. İnsanlar arasında rəqəmsallaşmanın sürətlə artması və mədəniyyətlərarası qarşılıqlı əlaqənin genişlənməsi müxtəlif cəmiyyətlərdə audiovizual məhsulların geniş şəkildə istehlak edilməsinə səbəb olmuşdur. Dil öyrənənlər üçün dinləmə və oxuma təkrarına olan ehtiyac da ani tərcümə tələbatını ortaya çıxarmışdır. Audiovizual materiallar üçün avtomatik altyazı tərcüməsinin inkişafı ilə insanlar artıq ani tərcümə ehtiyaclarını sürətli və səmərəli şəkildə qarşılıya bilirlər.

Bununla belə, avtomatik tərcümə sistemlərinin sürətli şəkildə yaratdığı çıxışların nə dərəcədə dəqiq olduğu və insan tərcüməsi ilə nə dərəcədə ekvivalentlik təşkil etdiyi hələ də tam müəyyən deyil. Maşın tərcüməsi sistemlərinin insan tərcüməçilərin xüsusi diqqət yetirdiyi mədəni ifadələri, emosional tonları və idiomatik istifadələri effektiv şəkildə nəzərə alıb-almadığı hələ də müzakirə mövzudur. Mədəniyyət daxilində formalaşan və dilə xas mənalara daşıyan sözlər, xüsusilə struktur və mədəni baxımdan bir-birindən uzaq dil cütlükləri arasında transferi çətinləşdirən elementlər kimi qiymətləndirilir (Benmessaoud et al., 2025). Buna görə də insan tərcümələri ilə maşın tərcümələrinin müqayisə edilməsi zərurətə çevrilmişdir. Avtomatik tərcümələr yüksək sürətlə hazırlandığı üçün qiymətləndirmə prosesinin də sürətli olması tələb olunur. Bu səbəbdən avtomatik keyfiyyət qiymətləndirmə metrikləri inkişaf etdirilmişdir.

Bu metriklər iki kateqoriyaya ayrılır. Ənənəvi metriklər əsasən n-qram əsaslıdır və səthi səviyyədə ekvivalentlik ölçümü aparır; buna görə də ənənəvi metriklər adlandırılır. Müasir metriklər isə qabaqcıl təbii dil emalı metodlarına əsaslanır və cümlə səviyyəsində semantik oxşarlığı və kontekstual əlaqələri qiymətləndirir. Bu metriklər insan mühakiməsi ilə uyğunlaşan daha əhatəli qiymətləndirmələr təqdim edir. Bu tədqiqatın təhlil bölməsində səthi qiymətləndirmədən daha dərin və insan yönümlü qiymətləndirmə imkanı verdiyi üçün müasir metriklərə üstünlük verilmişdir.

Təhlil prosesində bütün müasir metriklər tətbiq edilmişdir. BERTScore metriyi semantik oxşarlığı qiymətləndirir; LaBSE semantik ekvivalentliyi ölçür; COMET isə insan mühakiməsinə yönəlmiş semantik və dərin təhlil əsasında nəticələr yaradır. Tək bir metrik əsasında aparılan ölçmələr tərcüməni yalnız bir perspektivdən qiymətləndirir və məhdud nəticə təqdim edir. Bütün metriklərin birlikdə istifadə edilməsi ilə tərcümə səhvləri, semantik itkilər və kontekstual uyğunsuzluqlar kimi müxtəlif aspektlər daha əhatəli şəkildə qiymətləndirilir və daha etibarlı tərcümə keyfiyyəti nəticəsi əldə edilir.

Ənənəvi metriklər

BLEU

Papineni et al. (2002) tərəfindən inkişaf etdirilən BLEU dil müstəqilliyi, hesablanma asanlığı və insan mühakimələri ilə orta səviyyədə korrelyasiya göstərməsi səbəbindən geniş şəkildə istifadə olunmuşdur. Bu metrik tokenləşdirmə, qısalıq cəzası və n-qram uyğunluğu kimi bir sıra parametrlər əsasında qiymətləndirmə aparır (Kuwa et al., 2018; Post, 2018). Lakin onun bir neçə parametrdən asılı olması nəticələrin ardıcılığını tez-tez azaldır və müxtəlif tədqiqatlar arasında müqayisə edilə bilməyən skorların ortaya çıxmasına səbəb olur. Bu məhdudiyyətləri aradan qaldırmaq və eyni şərtlər altında təkrarlana bilən qiymətləndirməni təmin etmək məqsədilə tədqiqatçılar BLEU-nun standartlaşdırılmış versiyası kimi SacreBLEU modelini təqdim etmişlər (Post, 2018). Statistik maşın tərcüməsi üzərində çalışan tədqiqatçılar BLEU-dan model etibarlılığını yoxlamaq üçün tez-tez istifadə edirlər; çünki bu metrik avtomatik və xərc baxımından səmərəlidir (Tran et al., 2019). Bununla belə, BLEU əsasən səthi leksik uyğunluğa əsaslandığı üçün morfoloji baxımdan mürəkkəb və ya terminoloji cəhətdən zəngin dillərdə məhdud nəticələr göstərə bilər.

ROUGE

Mətn xülasələşdirməsinin keyfiyyətini qiymətləndirmək üçün 2004-cü ildə inkişaf etdirilən ROUGE, maşın tərəfindən yaradılan xülasələrlə insan tərəfindən yazılmış xülasələr arasındakı n-qram oxşarlığını ölçən, xatırlama əsaslı bir metrikdir (Steffes et al., 2023). ROUGE avtomatik sistemlər tərəfindən hazırlanan xülasələrlə insan tərəfindən yazılmış xülasələr arasındakı n-qram örtüşməsinə araşdırır (Ganesan, 2018). Liu və Yang (2010) tərəfindən yayımlanan məqalədə bu xülasələrin insan mühakimələri ilə yüksək korrelyasiya göstərdiyi bildirilmişdir. Bununla belə, ROUGE xülasələşdirmə qiymətləndirməsində standart metrik kimi qəbul edilsə də, sinonim sözləri və mövzu əhatəliliyini tanımaqda məhdud qaldığı üçün xülasələrin həqiqi keyfiyyətini tam əks etdirə bilmir (Ganesan, 2018). Buna görə də yalnız səthi səviyyədə oxşarlıq təqdim edir (Steffes et al., 2023).

METEOR

METEOR metriyi ilk dəfə Lavie və həmkarları tərəfindən 2004-cü ildə irəli sürülmüş, 2005-ci ildə Banerjee və Lavie tərəfindən inkişaf etdirilmiş, daha sonra isə 2007-ci ildə Lavie və Agarwal tərəfindən yenidən işlənmişdir (Lavie & Agarwal, 2007). BLEU-nun n-qram əsaslı oxşarlıq təhlilindəki məhdudiyyətlərini aradan qaldırmaq məqsədilə inkişaf etdirilən METEOR yalnız söz səviyyəsində dəqiqlik göstəricilərini deyil, həm də xatırlama dəyərlərini nəzərə alır. Bu metrik kök uyğunluğu, sinonim tanıma və parafraz uyğunlaşdırmasını da əhatə edən çevik və struktur baxımından sərbəst qiymətləndirmə yanaşması təqdim edir. Uyğunlaşdırmalar əsasında dəqiqlik, xatırlama və onların harmonik ortalaması olan F-skör hesablanır. Kök və sinonim kimi morfoloji, leksik və semantik variasiyaları tanımaqla METEOR əvvəlki modellərlə müqayisədə səthi səviyyəli semantik oxşarlığı daha effektiv şəkildə müəyyən edə bilir (Lee et al., 2023).

Müasir metriklər

BERTScore

Zhang et al. (2020) tərəfindən inkişaf etdirilən BERTScore tərcümə çıxışlarının dəqiqliyini və ümumi keyfiyyətini qiymətləndirmək üçün hazırlanmış semantik qiymətləndirmə metriyidir. Bu metrik namizəd və istinad cümlələri arasındakı oxşarlığı onların kontekstual vektor təmsillərini müqayisə etməklə ölçür (Tang, Yousuf, & Jin, 2024). BERTScore-un nəzəri əsası Goldberg-in araşdırması ilə dəstəklənir; həmin araşdırmada BERT modelinin neyron şəbəkə kimi həm semantik, həm də qrammatik təhlillərdə uğurlu nəticələr göstərdiyi müəyyən edilmişdir (Goldberg, 2019).

BERTScore-un semantik səviyyədə istifadə etdiyi oxşarlıq ölçmə metodu kosinus oxşarlığıdır. Açıq gözlü uyğunlaşdırma texnikasından istifadə etməklə token səviyyəsində uyğunlaşdırmalar aparılır və dəqiqlik, xatırlama və onların harmonik ortalaması olan F1 skörləri hesablanır. Ənənəvi metriklərdən fərqli olaraq, 104 dili dəstəkləyən bu modelin insan qiymətləndirmələri ilə yüksək korrelyasiyaya malik olduğu qəbul edilir və o, müasir ədəbiyyatda ön plana çıxan metriklərdən biri hesab olunur (Zhang et al., 2020).

LaBSE

Google AI tədqiqatçıları tərəfindən 2020-ci ildə inkişaf etdirilən LaBSE (Language-agnostic BERT Sentence Embedding) dil-müstəqil cümlə vektorlaşdırma modelidir (Feng et al., 2022). Bu metriyin məqsədi çoxdilli təbii dil emalı sistemləri üçün yüksək keyfiyyətli, dil-müstəqil cümlə vektorları yaratmaqdır.

LaBSE BERT arxitekturası üzərində qurulmuş və cümlə səviyyəsində semantik təmsillər yaratmaq üçün hazırlanmış bir modeldir. Model ümumi cümlə mənasını CLS pooling və sıx qatlar vasitəsilə emal edir; bu qatlar əhatəli kontekstual vektor təmsilləri çıxarır. Hər bir dil üçün ayrı təmsil məkanları yaratmaq əvəzinə, LaBSE yüzdən artıq dildə paralel cümlələri ortaq semantik məkanda uyğunlaşdırır. Bu çoxdilli uyğunlaşdırma kontrastiv tərcümə sıralaması məqsədi vasitəsilə həyata keçirilir və bu da

müxtəlif dillərdə semantik baxımdan ekvivalent cümlələrin oxşar vektor təmsilləri əldə etməsinə imkan verir (Krasner et al., 2025).

Ədəbiyyatda aparılan tədqiqatlar LaBSE-nin müxtəlif dil cütlüklərində insan mühakimələri ilə yüksək korrelyasiya əldə etdiyini göstərir və bu xüsusiyyət onu ən etibarlı müasir qiymətləndirmə metriklərindən biri kimi ön plana çıxarır (Feng et al., 2022).

COMET

2020-ci ildə Rei et al. tərəfindən inkişaf etdirilən COMET, avtomatik tərcümə çıxışları ilə insan qiymətləndirmələri arasında yüksək korrelyasiya əldə etməyi hədəfləyən müasir qiymətləndirmə metriyidir (Rei et al., 2020). Bu metrik mənbə cümlə, hipotez tərcümə və istinad tərcümədən ibarət üçlü giriş strukturunu emal edən dərin öyrənmə modeli vasitəsilə skorlar yaradır.

COMET modelinin insan mühakimələri ilə yüksək korrelyasiya göstərməsi onun tərcümə keyfiyyətinin qiymətləndirilməsi üçün xüsusi olaraq uyğunlaşdırılmış, əvvəlcədən öyrədilmiş çoxdillə modelin incə tənzimlənmiş versiyası olmasından irəli gəlir (Zouhar et al., 2024). Ənənəvi metriklər kimi yalnız səthi söz uyğunluğuna əsaslanmaq əvəzinə, COMET tərcümə keyfiyyətini semantik adekvatlıq, axıcılıq və mənanın qorunması kimi meyarlar əsasında qiymətləndirir. İnsan mühakimələri ilə yüksək korrelyasiya göstərən COMET modeli yalnız cümlə səviyyəsində qiymətləndirmə aparmaqla kifayətlənməmiş, eyni zamanda tərcümə səhvlərini müəyyən etmək və daha ətraflı təhlil təqdim etmək üçün də inkişaf etdirilmişdir. Bu təkmilləşdirmələr nəticəsində xCOMET kimi genişləndirilmiş variantlar təqdim edilmişdir (Guerreiro et al., 2024).

Metodlar

Bu tədqiqat ərəb–türk dil cütlüyündə insan tərcümələri ilə avtomatik tərcümələri təhlil etmək üçün müqayisəli tədqiqat dizaynından istifadə edir. Məqalənin ümumi həcmi və əhatə dairəsi nəzərə alınaraq Language Reactor əlavəsi vasitəsilə Netflix serial və filmlərindən beş audiovizual nümunə seçilmişdir. Həm insan tərəfindən hazırlanmış altı yazılar, həm də sistem tərəfindən yaradılmış altı yazılar insan qiymətləndirməsi ilə güclü uyğunluq göstərən üç müasir metrik — BERTScore, LaBSE və COMET — vasitəsilə qiymətləndirilmişdir. Hər bir tərcümə cütü eyni şərtlər altında müstəqil şəkildə araşdırılmışdır. Təhlilin əsas məqsədi insan tərcümələri ilə maşın tərcümələri arasındakı semantik, struktur və kontekstual uyğunluğu müəyyən etməkdir. Bu tədqiqatda bütün metrik əsaslı təhlillər Google Colab mühitində, BERTScore, LaBSE və COMET modellərinin açıq mənbəli tətbiqlərindən istifadə edilməklə aparılmışdır. Qiymətləndirmə prosesi tam avtomatlaşdırılmış və obyektiv şəkildə həyata keçirilmiş, əldə edilən bütün nəticələr heç bir manuel müdaxilə və ya dəyişiklik edilmədən olduğu kimi təqdim edilmişdir.

Nəticələr və müzakirə

Nümunə 1

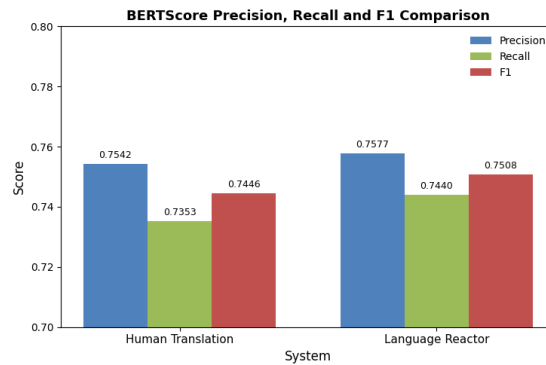
Növ	Tərcümə (ərəb dili)
İnsan tərcüməsi	الحياة مجرد لحظة. يبدأ كل شيء بلحظة.
Language Reactor tərcüməsi	الحياة مجرد لحظة. كل شيء يبدأ بلحظة.

Mənbə. Hayat dediğın sadəcə bir an. Her şey bir anla başlar (Midnight at the Pera Palace, 2022).

Mənbə mətn və istinad tərcümələri nəzərdən keçirildikdə hər iki ərəb dili tərcüməsinin orijinal mətnin əsas mənasını dəqiq və paralel şəkildə ötürdüyü görünür. İnsan tərcüməsi mənbə mətnin mənasına

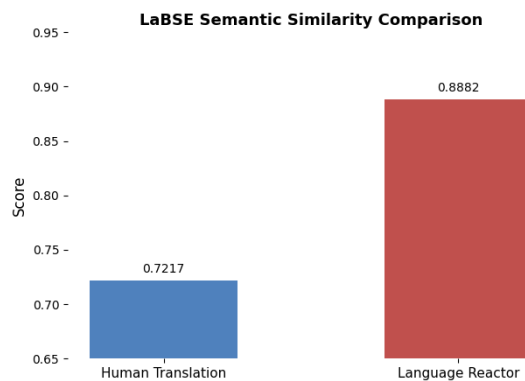
sadiq qalmaqla yanaşı, daha özünəməxsus sintaktik quruluş təqdim edir; bu isə axıcılığı və təbiiliyi artırır. Buna qarşılıq olaraq, Language Reactor tərcüməsi həm mənanı, həm də orijinal mətnin quruluşunu qoruyur və leksik vahidlərin ardıcılığını daha birbaşa şəkildə təkrarlayır. Bu, düzgün tərcümə nəticəsi yaratsa da, insan tərcüməsi ilə müqayisədə daha mexaniki görünür. Cümlədə mədəni referans və ya məcazi ifadə olmadığı üçün hər iki tərcümə semantik itkiyə yol vermədən ekvivalent mənanı uğurla çatdırır.

BERTScore təhlili



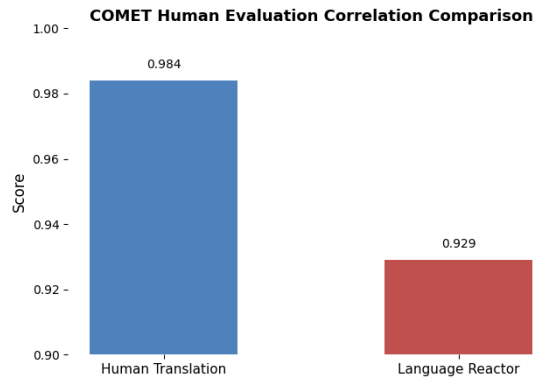
BERTScore təhlili nəzərdən keçirildikdə, Language Reactor (LR) tərcüməsinin insan tərcüməsindən (İT) daha yüksək nəticə əldə etdiyi görünür. Bu göstərici LR çıxışının mənbə mətnlə daha güclü semantik ekvivalentlik nümayiş etdirdiyini göstərir. LR və İT skorlarının bir-birinə yaxın olması isə hər iki tərcümənin bir-biri ilə yüksək dərəcədə uyğun nəticələr ortaya qoyduğunu göstərir. LR tərcüməsinin İT ilə müqayisədə daha yüksək skor əldə etməsi BERTScore-un iş prinsipi ilə əlaqələndirilə bilər. BERTScore daha yüksək semantik ekvivalentlik nümayiş etdirən tərcümləri müsbət qiymətləndirdiyi üçün bu nəticə avtomatik sistemin tərcümədə semantik və leksik uyğunluğu uğurla müəyyən edib əks etdirdiyini göstərir.

LaBSE təhlili



LaBSE təhlilinə görə, LR tərcüməsi 0.8882 skor əldə etmişdir və bu nəticə insan tərcüməsinin 0.7217 skorundan yüksəkdir. LaBSE cümlə səviyyəsində çoxdillə semantik uyğunluğu ölçüdü üçün LR tərcüməsinin daha yüksək nəticə əldə etməsi onun qiymətləndirilən semantik təmsillə daha güclü uyğunluq nümayiş etdirdiyini göstərir. İki skor arasındakı fərqin nisbətən məhdud olması və nəticələrin 1.0-a yaxınlaşması hər iki tərcümənin ümumi baxımdan semantik uyğunluq nümayiş etdirdiyini göstərir. Bununla yanaşı, LR tərcüməsinin daha yüksək nəticəsi onun mənbə mətnə daha yaxın struktur və semantik düzülüş təqdim etməsi ilə izah oluna bilər.

COMET təhlili



COMET metriyi BERTScore və LaBSE ilə müqayisədə fərqli nəticə ortaya qoymuşdur. İnsan mühakimələri əsasında öyrədilmiş neyron metrik kimi COMET insan tərcüməsini LR tərcüməsindən daha yüksək qiymətləndirmişdir. İki nəticə arasındakı fərqin çox böyük olmaması isə bu nümunədə LR tərcüməsinin insan tərcüməsinə yaxın performans göstərdiyini ortaya qoyur.

Tərcümə nümunələri nəzərdən keçirildikdə, hər iki çıxışın leksik seçim baxımından demək olar ki, eyni olduğu, yalnız söz sırası baxımından müəyyən fərqlər göstərdiyi müşahidə edilir. Bu nümunə COMET-in sadəcə səthi səviyyəli bir metrik olmadığını, əksinə semantik ekvivalentliyi və hədəf dil axıcılığını qiymətləndirərək insan yönümlü qiymətləndirmə meyarlarını skorlaşdırma prosesinə daxil etdiyini göstərir.

Nümunə 2

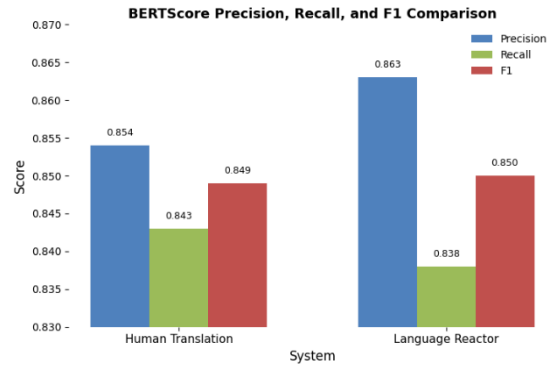
Növ	Tərcümə (ərəb dili)
İnsan tərcüməsi	يتطلب السعي وراء الأحلام الشجاعة.
Language Reactor tərcüməsi	يتطلب الأمر شجاعة لتحقيق الأحلام.

Mənbə. Hayallerin peşindən gitmek cesaret ister (My Brother 2, 2017).

Mənbə mətnində insanın məqsədlərinin ardınca qətiyyətlə getməsinə və arzularını həyata keçirmək üçün səy göstərməsini ifadə edən idiomatik bir ifadə yer alır. Hər iki tərcümə əsas semantik mənanı uğurla çatdırırsa da, dil təbiiliyi və üslub seçimi baxımından bir-birindən fərqlənir. İnsan tərcüməsi idiomatik ifadəni وراء الأحلام السعي şəklində verir; bu ifadə “arzuların ardınca getmək” mənasına daha yaxındır və gündəlik ərəb dili istifadəsinə uyğun, təbii və axıcı bir quruluş təqdim edir.

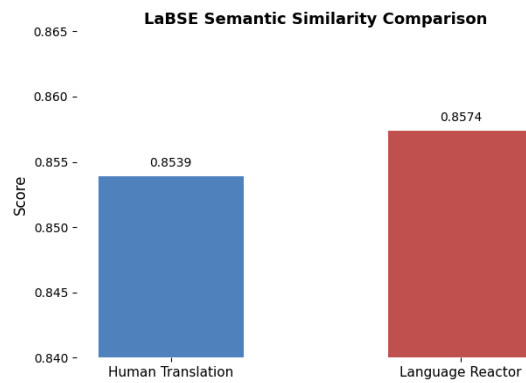
Buna qarşılıq olaraq, Language Reactor tərcüməsi eyni semantik məzmunu يتطلب الأمر شجاعة لتحقيق الأحلام ifadəsi ilə çatdırır. Bu isə arzuların ardınca getməkdən daha çox onların həyata keçirilməsini vurğulayan, izahlı və təsviri bir ifadə kimi görünür. Semantik baxımdan düzgün olsa da, bu quruluş daha rəsmi və idiomatiklik baxımından nisbətən zəif təsir bağışlayır. Hər iki tərcümə əsas mesajı qoruyur; bununla belə, insan tərcüməsi hədəf dildə daha yüksək axıcılıq və idiomatik dəqiqlik nümayiş etdirir.

BERTScore təhlili



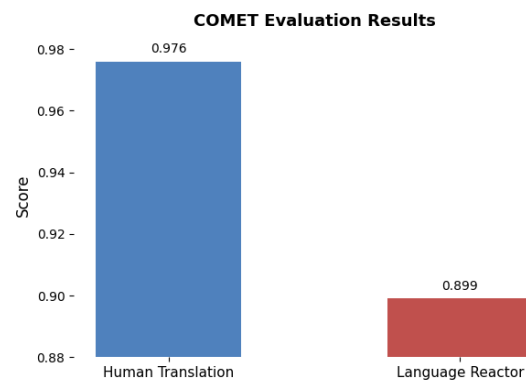
BERTScore təhlilinə görə, hər iki tərcümə 1.0-a yaxın skorlar əldə etmişdir. Bu nəticə onların mənbə mətnin mənasını dəqiq şəkildə çatdırdığını göstərir. Xüsusilə dəqiqlik və xatırlama göstəricilərinin 112emantic ortalamasını ifadə edən F1 dəyəri Language Reactor (LR) tərcüməsinin mənbə mətnin leksik düzülüşünü və kontekstual uyğunluğunu insan tərcüməsi ilə müqayisədə bir qədər daha yaxşı qoruduğunu göstərir.

LaBSE təhlili



LaBSE skor təhlilinə görə, həm insan tərcüməsi, həm də Language Reactor (LR) tərcüməsi bir-birinə çox yaxın və 1.0-a yaxın skorlar əldə etmişdir. Bu nəticə hər iki tərcümənin mənbə mətnlə yüksək səviyyədə 112emantic oxşarlıq nümayiş etdirdiyini göstərir. Bununla belə, LR tərcüməsinin bir qədər yüksək skor alması onun 112emantic ekvivalentlik və leksik uyğunluq baxımından mənbə mətnə daha yaxın bir versiya yaratdığını göstərir.

COMET təhlili



COMET skor təhlilinə görə, insan tərcüməsi (İT) Language Reactor (LR) tərcüməsi ilə müqayisədə daha yüksək keyfiyyətli çıxış təqdim etmişdir. Hər iki tərcümənin 1.0-a yaxın yüksək skorlar əldə etməsi onların ümumi performansının güclü olduğunu göstərir. Bununla belə, COMET 113emantic uyğunluğu və kontekstual adekvatlığı qiymətləndirdiyi üçün İT skorunun daha yüksək olması insan tərcüməsinin mənbə mətnlə məna uyğunluğu baxımından daha güclü ardıcılıq və axıcılıq təmin etdiyini göstərir.

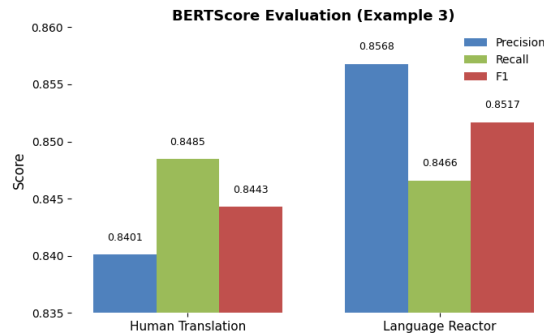
Nümunə 3

Növ	Tərcümə (ərəb dili)
İnsan tərcüməsi	كما أصدر الخبراء تحذيرا من هبوب عاصفة الليلية.
Language Reactor tərcüməsi	كما يحذر الخبراء من العواصف في المساء.

Mənbə. Uzmanlar, akşam saatleri için de fırtına uyarısı yapıyor
(Midnight at the Pera Palace, 2022).

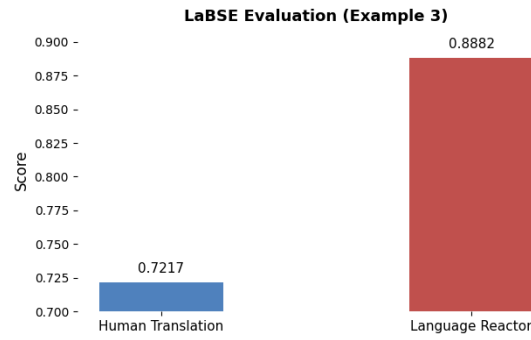
Mənbə mətn nəzərdən keçirildikdə dialoqun xəbər diskursuna xas məlumatlandırıcı bir tona malik olduğu müşahidə edilir. Üçüncü nümunə əsasında hər iki tərcümənin mənanı dəqiq və ardıcıl şəkildə çatdırdığı görünür. İnsan tərcüməsi ifadəni formal dəqiqliklə təqdim edir və xüsusilə zaman ifadəsi — “axşam saatları” — və xəbərdarlıq bildirən “fırtına xəbərdarlığı” kimi elementlərdə xəbər dili üslubunun normalarına daha yaxın qalır. Language Reactor (LR) tərcüməsi də düzgün şərh təqdim edir; lakin bir qədər fərqli leksik seçimlərdən istifadə etdiyi üçün semantik baxımdan doğru olsa da, xəbər mətninin kontekstində insan tərcüməsi qədər axıcı və təbii görünür.

BERTScore təhlili



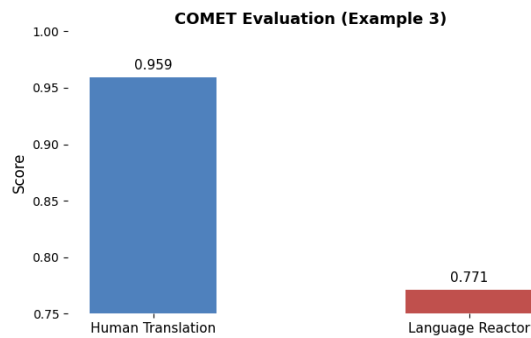
BERTScore nəticələri nəzərdən keçirildikdə insan tərcüməsi (İT) və Language Reactor (LR) skorlarının bir-birinə çox yaxın olduğu müşahidə edilir. Bununla belə, kiçik fərqlə LR tərcüməsi mənbə mətnlə daha yüksək uyğunluq nümayiş etdirir. Bu nəticə LR çıxışının İT ilə müqayisədə daha güclü leksik və kontekstual ekvivalentlik təqdim etdiyini göstərir.

LaBSE təhlili



LaBSE nəticələrində LR tərcüməsi İT-dən daha yüksək skor əldə etmişdir. LR-in daha yüksək nəticə göstərməsi onun tərcümə cümləsinin mənbə mətnə 114emantic baxımdan daha yaxın olduğunu göstərir. LR çıxışı maşın tərcüməsi kimi həm mənanı, həm də mənbə mətnə yaxın sintaktik quruluşu qoruduğu üçün bu vəziyyət LaBSE qiymətləndirməsində daha yüksək ümumi oxşarlıq skoruna səbəb olmuşdur.

COMET təhlili



COMET nəticələrində İT 0.959, LR isə 0.771 skor əldə etmişdir. COMET tərcümə keyfiyyətini 114emantic adekvatlıq, kontekstual uyğunluq və hədəf dil təbiiliyi əsasında qiymətləndirdiyi üçün İT skorunun daha yüksək olması insan tərcüməsinin daha təbii və ardıcıl quruluşa malik olduğunu göstərir. İT mənbə mətnlə tam səthi uyğunluq nümayiş etdirməsə də, onun axıcı və təbii ifadə quruluşu daha yüksək skorla nəticələnmişdir. LR çıxışı isə daha mexaniki təsir bağışladığı üçün müqayisədə daha aşağı nəticə əldə etmişdir.

Nümunə 4

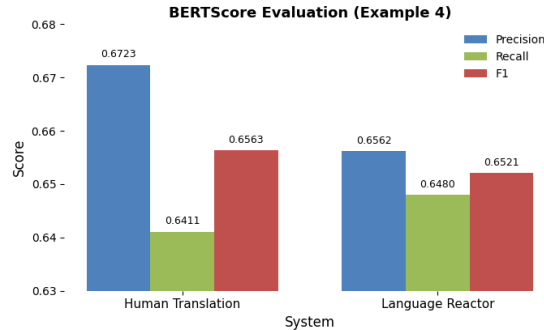
Növ	Tərcümə (ərəb dili)
İnsan tərcüməsi	توقفت عن الاكتراث لأراء الآخرين منذ زمن.
Language Reactor tərcüməsi	لقد مر وقت طويل منذ أن توقفت عن الاهتمام بما يعتقده الآخرون.

Mənbə. Başkalarının ne düşündüğüyle uğraşmayı bırakalı çok oldu (Another Self, 2022).

Mənbə mətn təhlil edildikdə cümlənin subyektiv ton və emosional dərinlik daşdığı müşahidə edilir. Hər iki tərcümə nəzərdə tutulan mənanı hədəf dilə dəqiq şəkildə ötürür. Bununla belə, İT cümləni daha birbaşa və minimalist şəkildə təqdim edir və mənbə mətnin emosional çalarını qoruyur. Buna qarşılıq olaraq, LR tərcüməsi mənbə mətndəki “çok oldu” ifadəsini لقد مر وقت طويل şəklinə daha açıq

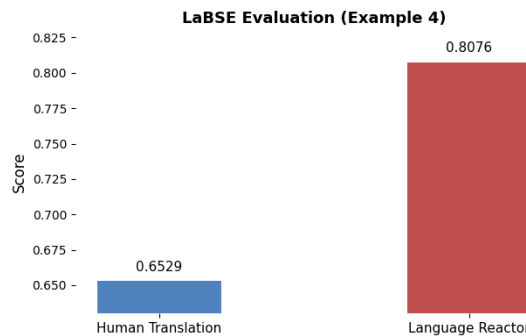
və zaman vurğusu güclü bir quruluşla təqdim edir. Bu seçim mənbə mətnin ümumi mənasını qorusa da, cümləni daha təsviri hala gətirir və insan tərcüməsinə nisbətən daha uzun bir ifadə yaradır.

BERTScore təhlili



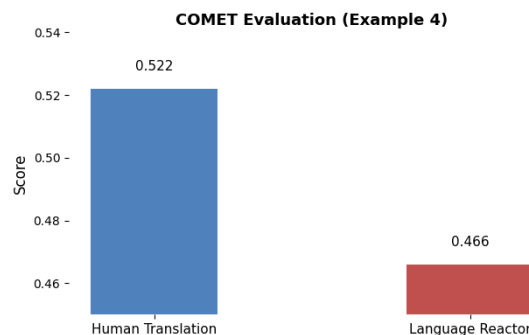
BERTScore nəticələrinə görə, İT LR ilə müqayisədə daha yüksək skor əldə etmişdir. Hər iki tərcümənin skorları bir-birinə yaxın olsa da, İT-nin daha yüksək nəticə göstərməsi onun mənbə mətnə semantik baxımdan daha yaxın olduğunu göstərir. LR tərcüməsi bir qədər aşağı skor əldə etsə də, BERTScore qiymətləndirməsinə görə hələ də qəbul edilə bilən və etibarlı bir çıxış təqdim edir.

LaBSE təhlili



LaBSE nəticələrinə görə, İT 0.6529, LR isə 0.8076 skor əldə etmişdir. Bu nəticə LR tərcüməsinin mənbə mətnlə daha güclü cümlə səviyyəli semantik uyğunluq nümayiş etdirdiyini göstərir. Bu nümunədə İT-nin daha aşağı skor alması onun daha yığcam və təbii ifadə quruluşu ilə bağlı ola bilər; belə bir yenidən ifadə etmə LaBSE qiymətləndirməsində semantik örtüşmənin daha aşağı görünməsinə səbəb olmuşdur.

COMET təhlili



COMET nəticələrinə görə, tərcümələrin heç biri xüsusilə yüksək skor əldə etməmişdir. İT-nin LR ilə müqayisədə bir qədər yüksək nəticə göstərməsi onun mənbə mətnə daha yaxın və hədəf dil baxımından daha təbii şəkildə qurulduğunu göstərir. Bununla belə, hər iki tərcümə çıxışının nisbətən aşağı skor alması onların orijinal kontekstdəki axıcılığı və təbiiliyi tam şəkildə qoruya bilmədiyini göstərir.

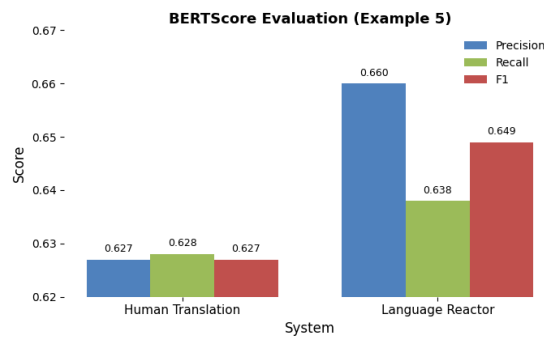
Nümunə 5

Növ	Tərcümə (ərəb dili)
İnsan tərcüməsi	ألن يزال من الممكن النهوض من ذلك الرماد؟
Language Reactor tərcüməsi	هل من الممكن أن تولد من جديد من هذا الرماد؟

Mənbə. Mümkün müdür o küllərdən yenidən doğmaq? (Another Self, 2022).

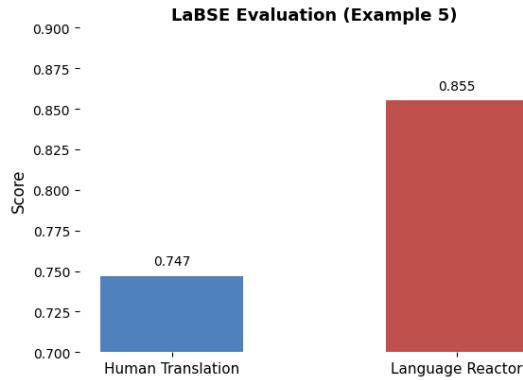
Mənbə mətn təhlil edildikdə “Mümkün müdür o küllərdən yenidən doğmaq?” ifadəsinin poetik və emosional yüklü bir ifadə olduğu görünür. Buradakı “yenidən doğmaq” ifadəsi metaforik xarakter daşıyır və sözün həqiqi mənasında doğulmanı deyil, yenilənmə, dirçəliş və özünü yenidən qurma mənasını ifadə edir. İnsan tərcüməsi “yenidən doğmaq” metaforunu النهوض ifadəsi ilə “yenidən ayağa qalxmaq” və ya “dirçəlmək” mənasına yaxınlaşdırır; bununla da mənbə mətnin poetik yenilənmə çalarını hədəf dildə daha təbii şəkildə yenidən qurur. Buna qarşılıq olaraq, Language Reactor tərcüməsi تولد من جديد ifadəsi ilə “yenidən doğulmaq” mənasını daha birbaşa şəkildə təqdim edir. Bu səbəbdən LR tərcüməsi mənbə ifadəyə leksik baxımdan daha yaxın görünsə də, insan tərcüməsi hədəf dildə daha təbii, emosional və poetik bir yenidənqurma təqdim edir.

BERTScore təhlili



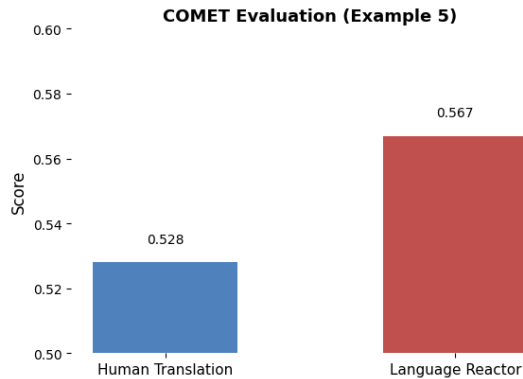
BERTScore təhlilinə görə, İT 0.627, LR isə 0.649 F1 skoru əldə etmişdir. Bu nəticələr hər iki tərcümənin mənbə mətnlə müqayisə edilə bilən səviyyədə semantik uyğunluq nümayiş etdirdiyini göstərir. Bununla belə, LR-in bir qədər yüksək skor əldə etməsi onun mənbə cümlənin mənasını semantik baxımdan daha yaxın şəkildə çatdırdığını göstərir. Hər iki tərcümə çıxışının oxşar nəticələr əldə etməsi onların struktur baxımından paralel olduğunu və semantik ekvivalentlik çərçivəsində etibarlı hesab edilə biləcəyini göstərir.

LaBSE təhlili



LaBSE nəticələrinə görə, İT 0.747, LR isə 0.855 skor əldə etmişdir. Semantik oxşarlığı ölçmək üçün hazırlanmış müasir bir metrik kimi LaBSE LR tərcüməsinə daha yüksək skor vermişdir; bu da onun mənbə mətnlə yüksək səviyyədə uyğunluq nümayiş etdirdiyini göstərir. İT-nin nisbətən aşağı skor əldə etməsi isə onun daha orijinal və yenidən qurulmuş ifadə formasının mənbə və hədəf cümlələr arasındakı semantik örtüşməni metrik baxımından bir qədər azaltması ilə izah oluna bilər.

COMET təhlili



COMET nəticələrinə görə, həm İT, həm də LR nisbətən yaxın skorlar əldə etmiş, LR isə bir qədər daha yüksək nəticə göstərmişdir. Məna dərinliyi, semantik adekvatlıq və dil uyğunluğu kimi meyarları nəzərə alan bir metrik kimi COMET bu nümunədə LR çıxışının mənbə mətnin semantik dərinliyini və nəzərdə tutulan mənasını İT ilə müqayisədə bir qədər daha effektiv şəkildə çatdırdığını göstərir.

Məhdudiyyətlər və gələcək tədqiqatlar

Bu tədqiqat ərəb–türk dil cütlüyünə əsaslanır və seçilmiş metriklərin performansını nümayiş etdirmək üçün kifayət edən təmsiledici nümunələr toplusunu əhatə edir. Gələcək tədqiqatlar nəticələrin etibarlılığını təmin etmək və daha da artırmaq məqsədilə fərqli dil cütlükləri və müxtəlif audiovizual tərcümə nümunələri üzərində aparıla bilər.

Nəticə

Müasir dövrdə texnologiyanın sürətli inkişafı və süni intellekt əsaslı vasitələrin geniş istifadəsi audiovizual məzmunun çoxdilli transferini xeyli asanlaşdırmış və sürətli tərcüməyə olan artan tələbatın daha səmərəli şəkildə qarşılmasına imkan yaratmışdır. Bu tədqiqat ərəb–türk dil

cütlüyündə insan tərcümələri ilə Language Reactor əlavəsi vasitəsilə yaradılan avtomatik maşın tərcümələrini müasir qiymətləndirmə metrikləri əsasında kəmiyyət baxımından müqayisə etmişdir.

Audiovizual tərcümələrin dil öyrənmə məqsədilə getdikcə daha geniş istifadə edilməsi, eləcə də ani və sürətli tərcüməyə olan ehtiyac bu sahənin məhdudiyətlərini aradan qaldırmağa yönəlmiş müxtəlif vasitələrin inkişafını təşviq etmişdir. Bu vasitələrdən biri olan Language Reactor əlavəsi istifadəçilərə audiovizual materiallardan altıyazı tərcümələrini ani şəkildə əldə etmək və onları insan tərcümələri ilə müqayisə etmək imkanı verir. Tədqiqatın analitik hissəsində audiovizual məhsullarda yer alan insan altıyazıları ilə Language Reactor çıxışları müqayisə edilmiş və qiymətləndirilmişdir.

Maşın tərcüməsinin inkişafı onun mədəni çalarları və dilin semantik dərinliyini insan tərcüməçilər qədər dəqiq və axıcı şəkildə ötürüb-ötürə bilməməsi ilə bağlı suallar doğurur. Bu səbəbdən maşın tərəfindən yaradılan çıxışların keyfiyyətini sürətli və etibarlı şəkildə qiymətləndirmək üçün müxtəlif qiymətləndirmə metrikləri inkişaf etdirilmişdir. N-qram əsaslı səthi oxşarlığa əsaslanan BLEU, ROUGE və METEOR kimi ənənəvi metriklər kontekstual əhatə dairələrinin məhdud olması səbəbindən bu tədqiqatın skorlaşdırma prosesində əsas qiymətləndirmə vasitəsi kimi istifadə edilməmişdir. Bunun əvəzinə semantik dəqiqlik, kontekstual ekvivalentlik və insan mühakiməsinə daha yaxın qiymətləndirmə imkanı təqdim edən BERTScore, LaBSE və COMET metriklərindən istifadə edilmişdir.

Təhlillər göstərmişdir ki, insan tərcümələri kontekstual dəqiqlik, axıcılıq və mədəni çalarların ötürülməsi baxımından Language Reactor tərcümələri ilə müqayisədə daha uğurlu nəticələr ortaya qoyur. Bununla belə, Language Reactor tərcümələri semantik uyğunluq və sintaktik ardıcılıq baxımından daha yüksək skorlar əldə etmişdir.

BERTScore nəticələrinə əsasən, Language Reactor tərcümələri beş nümunədən dördündə daha yüksək performans göstərmişdir. BERTScore token və alt-token səviyyəsində semantik ekvivalentliyi qiymətləndirdiyi üçün Language Reactor çıxışlarının yüksək skorları sistemin mənbə mətnə əsaslanan və struktur uyğunluğa yönəlmiş tərcümə quruluşu ilə əlaqələndirilə bilər. İnsan tərcümələri isə daha yaradıcı, təbii və mədəni baxımdan uyğunlaşdırılmış ifadələr təqdim etmişdir.

LaBSE təhlilləri Language Reactor-un bütün beş nümunədə daha yüksək nəticələr əldə etdiyini göstərmişdir. LaBSE cümlə səviyyəsində vektor təmsilləri vasitəsilə semantik ekvivalentliyi qiymətləndirdiyi üçün sistemin mənbə mətnə yaxın və uyğunlaşdırma yönümlü quruluşu BERTScore nəticələrinə bənzər şəkildə ardıcıl olaraq daha yüksək skorlarla nəticələnmişdir.

COMET təhlillərində insan tərcümələri ilə Language Reactor tərcümələri ümumilikdə müqayisə edilə bilən nəticələr ortaya qoymuşdur. Bununla yanaşı, insan tərcümələri bir sıra nümunələrdə daha yüksək axıcılıq və kontekstual uyğunluq nümayiş etdirmişdir. COMET insan mühakiməsinə əsaslanan qiymətləndirmə çərçivəsindən yararlandığı üçün insan tərcümələrinin semantik dərinlik və hədəf dil təbiiliyi baxımından üstün olduğu məqamları daha aydın şəkildə göstərmişdir.

Ümumilikdə, təhlillər insan və maşın tərcümələrinin müəyyən hallarda paralel və müqayisə edilə bilən nəticələr verdiyini göstərir. İnsan tərcümələri struktur uyğunluq baxımından bəzən mənbə mətnə uzaqlaşsa da, onların orijinallığı, mədəni uyğunlaşdırma qabiliyyəti və hədəf dildə təbiiliyi üstünlük təşkil edir. Bununla yanaşı, süni intellekt əsaslı tərcümə vasitələrinin sürətli inkişafı maşın tərcümələrinin keyfiyyətinin getdikcə insan tərcüməsinə daha yaxın nəticələr verə biləcəyini göstərir. Müəllifin məlumatına görə, Language Reactor çıxışlarının ərəb–türk audiovizual tərcüməsi kontekstində müasir qiymətləndirmə metrikləri vasitəsilə təhlilinə yönəlmiş tədqiqatlar hələ də məhduddur. Bu baxımdan hazırkı tədqiqat gələcək araşdırmalar üçün ilkin və istiqamətləndirici bir çərçivə təqdim edə bilər.

References

1. Ayık Akça, T. (2022). Edebi metinlerde ve uzmanlık alan metinlerinde makine çevirisinin olanakları/olanaksızlığı: Çevirmenin değişen görev tanımlarına yeniden bakmak. *RumeliDE Dil ve Edebiyat Araştırmaları Dergisi*, 30, 1321–1343. <https://doi.org/10.29000/rumelide.1188804>
2. Banitz, B. (2020). Machine translation: A critical look at the performance of rule-based and statistical machine translation. *Cadernos de Tradução*, 40(1), 54–71. <https://doi.org/10.5007/2175-7968.2020v40n1p54>
3. Barut, E. (2022). İstatistiksel makine çevirisi ile nöral makine çevirisinin dilbilimsel parametrelerle karşılaştırılması: Google Translate. *Akdeniz Havzası ve Afrika Medeniyetleri Dergisi*, 4(1), 103–118. <https://doi.org/10.54132/akaf.1116949>
4. Benmessaoud, S., Alzumar, F., & Abdulla, A. K. (2025). Preserving cultural difference in the translation of proverbs: A cultural linguistics approach. *Jordan Journal of Modern Languages and Literatures*, 17(3), 867–886. <https://doi.org/10.47012/jjml.17.3.6>
5. Brown, P. F., Cocke, J., Della Pietra, S. A., Della Pietra, V. J., Jelinek, F., Lafferty, J. D., Mercer, R. L., & Roossin, P. S. (1990). A statistical approach to machine translation. *Computational Linguistics*, 16(2), 79–85.
6. Díaz Cintas, J. (2012). Clearing the smoke to see the screen: Ideological manipulation in audiovisual translation. *Meta*, 57(2), 279–293.
7. Díaz Cintas, J., & Remael, A. (2021). *Subtitling: Concepts and practices*. Routledge.
8. Feng, F., Yang, Y., Cer, D., Arivazhagan, N., & Wang, W. (2022). Language-agnostic BERT sentence embedding. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 878–891). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2022.acl-long.62>
9. Ganesan, K. (2018). ROUGE 2.0: Updated and improved measures for evaluation of summarization tasks. *arXiv*. <https://doi.org/10.48550/arXiv.1803.01937>
10. Goldberg, Y. (2019). Assessing BERT's syntactic abilities. *arXiv*. <https://doi.org/10.48550/arXiv.1901.05287>
11. Göktürk, A. (1994). *Çeviri: Dillerin dili*. Yapı Kredi Yayınları.
12. Guerreiro, N. M., Rei, R., van Stigt, D., Coheur, L., Colombo, P., & Martins, A. F. T. (2024). xCOMET: Transparent machine translation evaluation through fine-grained error detection. *Transactions of the Association for Computational Linguistics*, 12, 979–995. <https://doi.org/10.48550/arXiv.2310.10482>
13. Karga Seven Pictures. (2022). *Midnight at the Pera Palace* [TV series]. Karga Seven Pictures.
14. Krasner, N., Vasselli, J., Ticona, B., Anastasopoulos, A., & Lo, C.-K. (2025). Machine translation metrics for Indigenous languages using fine-tuned semantic embeddings. In *Proceedings of the Fifth Workshop on NLP for Indigenous Languages of the Americas (AmericasNLP)* (pp. 100–104). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2025.americasnlp-1.11>
15. Kuwa, R., Kato, T., & Yamamoto, S. (2018). Classification of utterances based on multiple BLEU scores for translation-game-type CALL systems. *IEICE Transactions on Information and Systems*, E101-D(3), 750–757. <https://doi.org/10.1587/transinf.2017EDP7151>
16. Language Reactor. (n.d.). *Language Reactor*. <https://www.languagereactor.com/>
17. Lavie, A., & Agarwal, A. (2007). METEOR: An automatic metric for MT evaluation with high levels of correlation with human judgments. In *Proceedings of the Second Workshop on Statistical Machine Translation* (pp. 228–231). Association for Computational Linguistics.
18. Lee, S., Lee, J., Moon, H., Park, C., Seo, J., Eo, S., Koo, S., & Lim, H. (2023). A survey on evaluation metrics for machine translation. *Mathematics*, 11(4), Article 1006. <https://doi.org/10.3390/math11041006>
19. Lin, C.-Y. (2004). ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out* (pp. 74–81). Association for Computational Linguistics.

20. Liu, F., & Yang, L. (2010). Exploring correlation between ROUGE and human evaluation on meeting summaries. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(1), 187–196. <https://doi.org/10.1109/TASL.2009.2025096>
21. Maestri, A. (2021). Audiovisual translation at the service of the EU public communication 2.0. *Journal of Audiovisual Translation*, 4(3), 126–148. <https://doi.org/10.47476/jat.v4i3.2021.186>
22. OGM Pictures. (2022). *Another Self* (TV series). OGM Pictures.
23. Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). BLEU: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (pp. 311–318). Association for Computational Linguistics. <https://doi.org/10.3115/1073083.1073135>
24. Post, M. (2018). A call for clarity in reporting BLEU scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers* (pp. 186–191). Association for Computational Linguistics. <https://doi.org/10.18653/v1/W18-6319>
25. Ramizoğlu, R. (2024). Modern makine çevirisi teknolojilerine yönelik bir karşılaştırma. *Uluslararası Türkçe Edebiyat Kültür Eğitim (TEKE) Dergisi*, 13(1), 126–137. <https://doi.org/10.7884/teke.1317796>
26. Rei, R., Stewart, C., Farinha, A. C., & Lavie, A. (2020). COMET: A neural framework for MT evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2685–2702. Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.emnlp-main.213>
27. Sadıkov, T., & Sarıgül, K. (2021). Makine çeviri yöntemleri ve makine çevirisinin bugünkü durumu. *Uluslararası Türkçe Edebiyat Kültür Eğitim (TEKE) Dergisi*, 10(1), 192–205.
28. Slimani, S. (2025). Machine translation into Arabic: Challenges and alternatives. *Qədim Diyar Beynəlxalq Elmi Jurnal (Ancient Land International Scientific Journal)*, 7(2), 361–371. <https://doi.org/10.36719/2706-6185/44/361-371>
29. Steffes, B., Rataj, P., Burger, L., & Roth, L. (2023). On evaluating legal summaries with ROUGE. In *Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law* (pp. 457–461). ACM. <https://doi.org/10.1145/3594536.3595150>
30. TAFF Pictures. (2017). *My Brother 2* [Motion picture]. TAFF Pictures.
31. Tang, G., Yousuf, O., & Jin, Z. (2024). Improving BERTScore for machine translation evaluation through contrastive learning. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2024.3406993>
32. Tran, N., Tran, H., Nguyen, S., Nguyen, H., & Nguyen, T. N. (2019). Does BLEU score work for code migration? In *Proceedings of the 27th International Conference on Program Comprehension*. 165–176. IEEE. <https://doi.org/10.1109/ICPC.2019.00034>
33. Wang, H., Wu, H., He, Z., Huang, L., & Church, K. W. (2022). Progress in machine translation. *Engineering*, 18, 143–153. <https://doi.org/10.1016/j.eng.2021.03.023>
34. Yılmaz, S. (2023). Arapça–Türkçe çeviri türlerinde nöral makine çeviri modellerinin verimliliği: ChatGPT örneği. *Şarkiyat Mecmuası*, 43, 339–355. <https://doi.org/10.26650/jos.1324416>
35. Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2020). BERTScore: Evaluating text generation with BERT. In *Proceedings of the International Conference on Learning Representations*.
36. Zouhar, V., Chen, P., Lam, T. K., Moghe, N., & Haddow, B. (2024). Pitfalls and outlooks in using COMET. In *Proceedings of the Ninth Conference on Machine Translation*. 1272–1288. Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.wmt-1.121>