

Љиљана Тасић*

Универзитет у Нишу

Филозофски факултет

Докторанткиња

***АНАЛИЗА ПРЕВОДА ИДИОМА У ПРОГРАМУ ЗА
МАШИНСКО ПРЕВОЂЕЊЕ GOOGLE TRANSLATE НА
ПРИМЕРУ РОМАНА „ЧИТАВ ЖИВОТ“ И „ТРАФИКАНТ“
ЗА ЈЕЗИЧКИ ПАР НЕМАЧКО-СРПСКИ*****

Овај рад се бави анализом превода идиома са немачког на српски језик у програму за машинско превођење Google Translate. Корпус чине романи Роберта Зеталера „Читав живот“ и „Трафикант“. Циљеви истраживања били су (1) да се испита степен препознавања идиома у зависности од њиховог типа и (2) да се испитају потешкоће које доводе до грешака у преводу идиома у овом програму. Утврђено је да Google Translate боље препознаје потпуне и семи-непрозирне идиоме од семи-транспарентних идиома. Анализа грешака показала је да је најзаступљенија грешка у корпусу дословни превод идиома. Поред ове грешке идентификоване су и грешке дословног превода са енглеског језика, аномалије у форми и непреведени или делимично преведени идиоме. До ових грешака доводе различите потешкоће, попут непознавања идиома у тексту, посредног превода са енглеског на српски језик и лексичке и структурне некореспонденције између немачког и српског језика. Додатна анализа прецизних превода показала је да Google Translate у преводу идиома користи превод идиомам сличног значења и сличне или различите форме и парафразу.

Кључне речи: идиоме, машински превод, Google Translate, немачки језик, српски језик

* ljljana.tasic@filfak.ni.ac.rs

** Ово истраживање подржало је Министарство науке, технолошког развоја и иновација Републике Србије (Уговор бр. 451-03-47/2023-01/ 200165).

Увод

Машинско превођење једна је од најсложенијих области компјутерске лингвистике и означава превођење текста са једног на други природни језик помоћу компјутера (Рамлов и Кренц 2008: 20). Према Шванке (1991: 11) циљ програма за машинско превођење је да пружи преводе који су у квалитативном смислу у рангу са људским преводима. Том приликом се као посебан изазов издваја превод идиома. Истраживања која се баве евалуацијом машинског превода идиома углавном испитују прецизност у њиховом преводу (Базиотис и др. 2023; Шао и др. 2017; Боуамер и др. 2010; Салтон и др. 2014) или се баве анализом грешака (Коста и др. 2015; Манојловић и др. 2017; Тасић 2023). Недовољно прецизан превод и грешке приликом превођења идиома могу се јавити услед потешкоћа да се идиоми препознају у тексту, али и услед лексичких и структурних некореспонденција између језика, услед непостојања еквивалента у језику циљу и других разлога (Шторер и Швал 1995: 40–42; Бејкер 1992: 68–71). Како је препознавање идиома у тексту једна од највећих потешкоћа приликом машинског превода, у овом раду се испитује степен препознавања идиома у зависности од њиховог типа у програму за машинско превођење Google Translate. Такође, анализом грешака и испитивањем стратегија превођења испитују се све врсте потешкоћа које доводе до грешака у преводу идиома са немачког на српски језик. Истраживање је спроведено у програму за машинско превођење Google Translate, будући да је бесплатан и доступан на интернету. Google Translate спада у неуралне машинске преводиоце који функционишу помоћу неуралне мреже и вештачке интелигенције (Уану 2018: 881).

Идиоми

У литератури се могу пронаћи разни термини за означавање чврстих спојева речи. Са једне стране, у употреби су изрази: фразеологизам, фраза или фразема, док се са друге стране говори о идиомима (Флајшер 1982: 8). Према Флајшеру (1982: 34) фразеологизми представљају синтаксичке спојеве и они су надређени појам свим језичким јединицама које су предмет проучавања фразеологије. У њих се, међутим, не убрајају скупови речи као што су глаголски комплекси, рефлексивни глаголи и конструкције које се састоје од члана и именице. Фразеологизми морају да садрже макар једну аутосемантичку реч, чиме се такође искључује рекција глагола, односно спојеви попут *warten auf jemanden*. Да би се фразеологизми даље разграничили од слободних спојева речи, Флајшер (1982: 35) уводи следеће критеријуме:

идиоматичност, семантичко-синтаксичка стабилност, лексикализованост и репродуктивност. Идиоматичност представља нерегуларан однос између значења компоненти тог фразеологизма и значења целе реченице. Овај однос представљен је у другој реченици:

1. Gustav hat bei seinem Vater ein Auto in der Garage.

Густав има аутомобил код свог оца у гаражи.

2. Gustav hat bei seinem Vater einen Stein im Brett.

Густав је омиљен код свог оца.

Значење прве реченице изводи се на основу појединачних значења конституената ове реченице, док се значење друге реченице не може извести на основу појединачних значења речи Stein и Brett. Дакле, у овом случају реч је о фразеологизму bei jemandem einen Stein im Brett haben који поседује одређени степен идиоматичности, односно реч је о идиому. Критеријум идиоматичности уско је повезан са критеријумом стабилности. Замена компоненти фразеологизма са одређеним степеном идиоматичности нарушила би значење тог фразеологизма. Уколико би се у другој реченици израз im Brett заменио изразом in der Garage, целокупно значење идиома било би нарушено. Дакле, код овог фразеологизма постоји лексичко-семантичка стабилност. Критеријуми идиоматичности и стабилности доводе до лексикализовања фразеологизма. Реч је о похрањивању фразеологизма у лексикону. Приликом коришћења једне такве синтаксичке конструкције не долази до њене продукције већ се она као спремна лексичка јединица репродукује (Флајшер 1982: 35–41).

Према Бургеру (2010: 11–15), фразеологизми су чврсти спојеви који такође испуњавају одређене критеријуме. У оквиру фразеологије у ширем смислу, фразеологизми се састоје од најмање две речи и комбинација речи унутар фразеологизма је позната и употребљава се унутар једне језичке заједнице. Критеријум идиоматичности односи се на фразеологију у ужем смислу, а фразеологизми који га испуњавају представљају идиоме. Поред идиоматичности, Бургер говори и о критеријумима чврстоће и менталном похрањивању фразеологизма као јединица. Фернандо и Флавел (1981: 17) наводе пет критеријума којима описују идиоме. По угледу на претходне ауторе, наводи се да значење идиома није резултат слагања значења његових конституената. Идиом је целина која или има парњак у виду хомонима са дословним значењем или је макар један конституент употребљен дословно, иако се сам идиом не може дословно разумети. Такође, идиоми тешко подлежу трансформацијама, заједно чине скуп израза у једном језику и институционализовани су.

У литератури на српском језику у употреби је и термин лексички спој. Лексички спојеви настали су комбиновањем појединих лексема у веће, више или мање чврсте целине. Они представљају „уобичајено, редовно, а понекад и систематско јављање једне лексеме у спрези са најмање још једном унутар синтагме или реченице” (Прћић 2016, 147). Ради одређивања степена компактности лексичких спојева користе се три критеријума: заменљивост, постојаност и прозирност. Идиоми и колокације представљају два основна типа лексичких спојева. Заменљивост се односи на могућност замене лексема које се налазе унутар лексичког споја другим лексемама сродног значења. Постојаност је морфосинтаксичко својство, које подразумева могућност подвргавања целине трансформацијама или модификацијама. На крају, прозирност се односи на степен предвидљивости смисла целине. Уколико се лексички спој одликује могућом или минималном заменљивошћу, слабом постојаношћу и прозирношћу, реч је о колокацији. Уколико се пак лексички спој одликује немогућом или минималном заменљивошћу, високом постојаношћу и непрозирношћу, онда ће бити речи о идиому. Главно својство идиома јесте идиоматизованост. Сви идиоми су размештени дуж континуума прозирности и постојаности, те као такви показују виши или нижи степен непрозирности, али и могућности подвргавања трансформацијама или модификацијама (Прћић 2016: 147–148). Бејкер (1992: 63) такође дефинише идиоме поредећи их са колокацијама и наводи да су идиоми „замрзнути шаблони језика“ са мало или нимало могућности варијације конституената, као и да се значење идиома не може извести на основу појединачних речи у њему.

На основу представљених извора из литературе фразеологизми се могу сагледати као надређени појам другим фразеолошким јединицама. Идиоми су они фразеологизми који испуњавају критеријум идиоматичности, односно њихово значење се не може извести на основу појединачних значења конституената који их чине и одликују се одређеним степеном стабилности и устаљености унутар једне језичке заједнице. Такође, идиоми су распоређени дуж континуума прозирности, односно идиоматичности. Да би се идиоматичност, као главно својство идиома ближе одредила, неопходно је сагледати типове идиома.

Типови идиома

Постоји више начина како се идиоми могу разврстати. Једна од подела јесте према степену идиоматичности. Тако се идиом *bei jemandem einen Stein im Brett haben* одликује високим степеном идиоматичности, будући да

семантички однос између лексема Stein, Brett и haben није транспарентан. Флајшер (1982: 38) овакве идиоме назива потпуно идиоматичним. Следећу групу чине идиоми код којих је степен идиоматичности нижи и они се називају полуидиоматичним идиомима. У фразеологизму einen Streit vom Zaune brechen, именица Streit употребљена је у свом дословном значењу што доводи до нижег степена идиоматичности. Постоје изрази који се истовремено могу употребити и тумачити и као фразеологизми и као слободни спојев речи. Тако се израз auf der Straße liegen може разумети и у дословном значењу „лежати на улици”, али и у идиоматичном „бити незапослен”. Бургер (2010: 36) говори о фразеолошком и слободном значењу фразеологизама, између којих могу постојати различити односи. Што је већи несклад између фразеолошког и слободног значења, то ће степен идиоматичности бити на вишем ступњу. Услед тога разликују се идиоматични, полуидиоматични и неидиоматични фразеологизми.

Фернандо и Флавел (1981: 28) уводе и четврти тип фразеологизма у зависности од степена идиоматичности. Разликују се транспарентни изрази код којих се значење може извести из конституената и они заправо и нису идиоми. Следећи на скали идиоматичности јесу семи-транспарентне фразе које у исто време могу бити схваћене и у идиоматичном и дословном значењу, као што је идиом to add fuel to the fire. Потом следе семи-непрозирне фразе које представљају идиоме са значењем које није у потпуности неразумљиво, као што је идиом tarred with the same brush. На самом крају континуума налазе се непрозирне фразе или потпуни идиоми чије се значење не може наслутити на основу значења конституената, као у идиому to pass the buck.

Табела 1. Типови идиома

Идиом	bei jemandem einen Stein im Brett haben	einen Streit vom Zaune brechen	auf der Straße liegen	die Zähne putzen
Флајшер (1982)	потпуно идиоматични	полуидиоматични	Не наводи се специфичан назив (могу имати и дословно и идиоматично значење).	/
Бургер (2010)	идиоматични	полуидиоматични	/	неидиоматични
Фернандо и Флавел (1981)	непрозирни или потпуни идиоми	семи-непрозирне фразе	семи-транспарентне фразе	транспарентне фразе

Потешкоће приликом превођења идиома

Приликом превођења идиома преводиоци се сусрећу са различитим потешкоћама. Главну потешкоћу представља идентификација односно препознавање идиома, будући да фразеологизми са идиоматичним значењем нису увек лако уочљиви у тексту. Услед тога преводиоци лакше препознају идиоме који се теже разумеју у тексту, односно који у датом контексту немају много смисла. Са друге стране, спојеви који могу имати и дословно и идиоматично значење теже се идентификују као идиоми (Бејкер 1992: 65–67). У контексту машинског превођења Шторер и Швал (1995: 40) указују такође на проблематику превођења идиома који могу имати и дословно и идиоматично значење. Превођење оваквих идиома представља изазов, будући да је неопходно одредити њихово значење у датом контексту.

Поред препознавања идиома, Бејкер (1992: 68–72) наводи и друге потешкоће са којима се преводиоци сусрећу, попут непостојања еквивалентног идиома у циљном језику. Такође, одређено значење се у једном језику може исказивати помоћу једне речи, док се у другом језику исто значење може изражавати помоћу идиома. Идиоми могу бити и културолошки условљени, односно могу се односити на одређене друштвене или религијске прилике које не постоје у другим језицима. Потешкоћа може настати и када у циљном језику постоји сличан идиом, који се, међутим, користи у другачијем контексту. Тако су енглески идиом *to skate on thin ice* и српски идиом *навући некога на танак лед* слични по својој форми, као и по значењу, док су контексти у оквиру којих се ови идиоми користе различити. Енглески идиом означава непромишљено поступање које може бити опасно, а српски идиом односи се на довођење некога у опасну ситуацију.

Шторер и Швал (1995: 40–42) овакве потешкоће у машинском преводу разврставају на основу степена структурне и лексичке кореспонденције између идиома из изворног језика и његовог еквивалента у циљном језику. У наредним примерима представљени су различити степени еквиваленције који директно утичу на потешкоће приликом превођења.

Пример 1.

нем. *die Katze aus dem Sack lassen*

енгл. *to let the cat out of the bag*

срп. открити тајну

Пример 2.

нем. Zwei Fliegen mit einer Klappe schlagen

енгл. to hit two birds with one stone

срп. убити две муве једним ударцем

Пример 3.

нем. unter dem Pantoffel stehen

енгл. to be hen-pecked

срп. бити некеме подређен

Идиом на немачком језику из првог примера има еквивалент у енглеском како у структурном, тако и у лексичком погледу, услед чега је његово превођење на енглески језик олакшано. У другом примеру немачки идиом и његов енглески еквивалент деле исту унутрашњу структуру, али се лексички разликују. У немачком идиому употребљена је именица Fliegen (мушице), док је у енглеском употребљена именица birds (птице). Превођење је у овом случају отежано услед лексичке некореспонденције. У трећем примеру проблем настаје услед структурне некореспонденције која може настати између идиома изворног и циљног језика. Уколико би се немачки идиом употребио у реченици: Er stand unter dem Pantoffel seiner Chefin, проблем би настао приликом превођења на енглески језик, јер се енглески фразеологизам не може употребити на исти начин у реченици са именичком фразом „своје шефице“. Услед тога би се у преводу морала употребити парафраза.

Приликом превођења идиома од стране човека или машинског превода могу се јавити грешке до којих доводе различите потешкоће. Главну потешкоћу представља препознавање идиома у тексту, а посебно израза који могу имати и дословно и идиоматично значење и чији прецизан превод зависи од разумевања контекста. Поред културолошких разлика, на превод може утицати и степен лексичке и структурне кореспонденције између идиома изворног и циљног језика.

Истраживање

Циљеви овог истраживања били су да се испита (1) у којој мери програм за машинско превођење Google Translate препознаје идиоме у зависности од њиховог типа, односно од степена идиоматичности и (2) да се одреде потешкоће које доводе до грешака у преводу идиома. Ово истраживање треба да одговори на следећа истраживачка питања:

1. У којој мери Google Translate прецизно преводи идиоме у зависности од њиховог типа?
2. Које се грешке јављају приликом превођења идиома у зависности од њиховог типа?
3. Које стратегије превођења су коришћене код прецизних превода идиома у зависности од њиховог типа?

Први циљ истраживања испитује се најпре помоћу првог истраживачког питања, приликом чега се сви преводи идиома у корпусу означавају као прецизни или недовољно прецизни. Како прецизност у преводу не мора истовремено да значи и да је идиом у тексту препознат, додатно је спроведена анализа прецизних и недовољно прецизних превода. Помоћу другог истраживачког питања испитује се да ли се у преводу јавила грешка дословног превода, на основу које се може закључити да идиом није препознат у тексту. На крају се у трећем истраживачком питању проверавају стратегије превођења код прецизних превода. Код идиома који имају еквиваленте у форми и значењу на српском језику не може се са сигурношћу рећи да су прецизно преведени услед њиховог препознавања. Стога се проверава заступљеност других стратегија превођења са посебним освртом на идиоме у корпусу који немају еквиваленте у форми и значењу на српском језику. Други циљ истраживања испитује се помоћу другог истраживачког питања. На основу идентификованих грешака у корпусу утврђују се потешкоће које су довеле до тих грешака.

Корпус и методологија

Корпус чине два књижевна текста на немачком језику: романи „Читав живот” и „Графикант“ аутора Роберта Зеталера. Током истраживања коришћен је и превод ових романа на српски језик.

У првом кораку истраживања романи на немачком језику су прочитани и идиоми су кодирани. Корпус чине два поткорпуса, обележена великим словима ЧЖ и ТР. Сваком идентификованом примеру додељен је редни број према редоследу појављивања у роману. На тај начин је први идентификовани идиом у поткорпусу ЧЖ кодиран на следећи начин:

1. Abgemagert bis auf die Knochen [ЧЖ001]

У овом истраживању коришћен је термин идиом. Идентификација идиома извршена је на основу критеријума наведених у теоријском делу (Флајшер 1982; Бургер 2010; Фернандо и Флавел 1981; Прћић 2016). Реч је, дакле, о спојевима од најмање две речи који имају одређени степен идиоматично-

сти и стабилности. За разврставање идиома према степену идиоматичности, употребљена је класификација коју предлажу Фернандо и Флавел (1981: 28), будући да најпрецизније именује све типове идиоматичности. Разликују се, дакле, потпуни, семи-непрозирни, семи-транспарентни идиоми и транспарентне фразе. Дефиниције других аутора узете су такође у обзир. Код потпуних идиома значење се не може извести на основу његових појединачних конституената, са чим су сви аутори сагласни. Код семи-непрозирних идиома значење није у потпуности неразумљиво. Са овим типом могу се упоредити Флајшерови полуидиоматични идиоми код којих је један конституент употребљен у свом дословном значењу. Будући да је ФлајшEROVA дефиниција прецизнија, за одређивање овог типа користиће се такво сагледавање. Потом следе семи-транспарентни идиоми који могу имати и дословно и идиоматично значење у зависности од контекста. Транспарентне фразе нису сагледаване у овом истраживању, будући да нису идиоматичне. На основу ове класификације кодирани идиоми су разврстани према степену идиоматичности на потпуне, семи-непрозирне и семи-транспарентне идиоми.

У другом кораку преведени су одсечци текста који садрже идиоми са немачког на српски језик у програму Google Translate. Приликом превода се водило рачуна о томе да одсечак текста пружи одговарајући контекст за правилну интерпретацију идиома у њему. Потом је извршена квалитативна анализа током које је превод сваког идентификованог идиома у тексту оцењен као прецизан или недовољно прецизан. Прецизност је одређена на основу тога да ли превод представља неку од стратегија превођења идиома (идиом сличне форме и сличног значења, идиом сличног значења а различите форме, парафразу и изостављање) (Бејкер 1992), а којом се идиоматично значење преноси на исти начин у српском као и у немачком језику. Приликом оцењивања прецизности превода, коришћен је поменути референцијални превод романа на српски језик. Такође, случајни узорак од 40 преведених идиома из референцијалног превода упоређен је са преводом тих идиома из Српско-немачког преводног фразеолошког речника ради утврђивања прецизности и поузданости референцијалног превода.

У трећем кораку истраживања примењена је квалитативна анализа. Најпре је спроведена анализа грешака код недовољно прецизних превода идиома. Инструмент за анализу грешака чини таксономија сачињена за потребе овог истраживања. Манојловић и др. (2017) приликом анализе превода идиома за језички пар енглеско-хрватски, користе следећу таксономију: еквивалент, адекватно значење, дословни превод, аномалија у форми, непре-

ведено или делимично преведено, погрешан идиом и остало. За разлику од поменуте таксономије, овај рад одвојено сагледава прецизне и недовољно прецизне преводе идиома, те ће за анализу грешака бити размотрене само следеће грешке: дословни превод, аномалија у форми и непреведен или делимично преведен идиом. Превод погрешним идиомом није примећен у корпусу, те ни ова грешка није уврштена у таксономију. Док се код дословног превода идиом преводи као регуларан текст, аномалија у форми се односи на идиоме које је програм препознао и превео, али које имају неку врсту грешке попут погрешне речи у преводу. Тасић (2023) указује на подтип дословног превода која настаје услед дословног превода са енглеског на српски, иако је у питању превод са немачког на српски језик. Оваква грешка настаје услед посредног превођења у програмима за машинско превођење, које се одвија у два корака, и то са немачког на енглески и потом са енглеског на српски језик. У корпусу су стога идентификоване следеће врсте грешака: дословни превод, аномалија у форми, непреведен или делимично преведен идиом и превод са енглеског на српски језик. У истраживању је накнадно извршена и анализа прецизних превода идиома у циљу тачнијег одређивања потешкоћа.

Анализа

У истраживању је идентификовано 152 идиома, од чега 86 у роману „Читав живот“ и 66 у роману „Трафикант“. Прво истраживачко питање односи се на одређивање прецизности превода у зависности од типа идиома у програму за машинско превођење Google Translate. Одговор на прво истраживачко питање дато је у табели 2.

Табела 2. Прецизни и недовољно прецизни преводи идиома

	Прецизни преводи	Недовољно прецизни пре- води
Потпуни идиоми	48	32
Семи-непрозирни идиоми	25	15
Семи-транспарентни идиоми	12	20
Укупан број:	85	67

Из табеле 2 може се увидети да је већи број прецизних превода потпуних и семи-непрозирних идиома, док је код семи-транспарентних идиома већи број недовољно прецизних превода. У наставку су анализирани недовољно прецизни и прецизни преводи идиома.

Недовољно прецизни преводи идиома

У оквиру другог истраживачког питања извршена је анализа грешака недовољно прецизних превода идиома. У табели 3 представљен је однос ових грешака у зависности од типа идиоматичности.

Табела 3. Расподела грешака у корпусу

	Дословни превод	Превод са енглеског језика	Аномалија у форми	Непреведено/ делимично преведено
Потпуни идиоми	26	3	2	1
Семи-непрозирни идиоми	13	1	1	0
Семи-транспарентни идиоми	19	0	0	1
Укупан број:	58	4	3	2

На основу табеле 3 може се увидети да је дословни превод најчешћа грешка у корпусу за све типове идиома. На основу великог броја дословних превода увиђа се да је главна потешкоћа приликом превода идиома у програму Google Translate њихово препознавање у тексту. Ова грешка илустрована је у примерима 4-6. Превод програма Google Translate означен је скраћеницом ГТ, а референцијални превод као РП.

Пример 4.

seine Geschäfte erledigen [ЧЖ011]

ГТ: Обављати свој посао

РП: Обављати нужду

У примеру 4 представљен је семи-транспарентан идиом seine Geschäfte erledigen, који може имати и дословно значење обављати посао и идиоматично значење обављати нужду. На основу контекста увиђа се да је реч о идиоматичном значењу, које програм за превођење није препознао.

Пример 5.

Der Tod hockte ihm hinter der Stirn. [ЧЖ002]

ГТ: Смрт му чучи иза чела.

РП: Смрт му је за вратом.

У овом примеру приказан је семи-непрозирни идиом, будући да се именица смрт налази у свом дословном значењу. Грешка је и у овом примеру настала услед дословног превода осталих компоненти идиома, на основу чега се закључује да идиом није препознат.

Следећу идентификовану грешку у корпусу представља превод са енглеског језика и она је приказана у наредном примеру:

Пример 6.

Die Ahnungslosigkeit ist ja praktisch das Gebot der Stunde. [TP061]

ГТ: Незнање је ред дана.

РП: На цени је да ни о чему немаш појма.

Идиом на немачком језику *das Gebot der Stunde* препознат је и преведен на његов еквивалент на енглеском језику *the order of the day* који је даље преведен дословно на српски језик као ред дана. Потешкоћу представља, дакле, посредни превод са немачког на српски преко енглеског језика.

У корпусу је уочена и грешка аномалије у форми идиома у циљном језику. Реч је о идиомима који су препознати, али је у преводу примећена нека врста аномалије у форми, као у примеру 7.

Пример 7.

an den Haaren herbeigezogene Frage [TP033]

ГТ: натегнуто питање

РП: неосновано питање

С обзиром на то да идиом *an den Haaren herbeigezogen* није преведен дословно, увиђа се да је идиом у програму препознат. Са друге стране, у преводу програма Google Translate употребљен је неадекватан придев којим се идиоматично значење не преноси на исти начин као придевом у референцијалном преводу.

Последњу грешку представља непреведен или делимично преведен идиом. Ова грешка приказана је у примеру 8.

Пример 8.

Halt deine Goschn! [TP035]

ГТ: Држи свој Госцхн!

РП: Зачепа губицу!

Идиом на немачком језику преведен је делимично будући да именица *Goschn* није преведена на српски језик.

Као што је већ речено, велики број грешака у преводу чине дословни преводи. Анализа прецизних превода, која је приказана у наставку рада, показује је да дословни преводи могу бити оцењени и као прецизни преводи идиома. Детаљнијом анализом примера показано је да на прецизност у преводу идиома утиче лексичка и структурна кореспонденција између немачког и српског језика. Односно, иако програм не препознаје идиом, услед помнуте кореспонденције дословни превод у српском језику може преносити значење на исти начин као идиом у немачком језику. Пример 9 илуструје дословни превод који је означен као непрецизан, а на кога је утицала лексичка некореспонденција идиома у ова два језика.

Пример 9.

auf dem Lohnzettel der Firma stehen [ЧЖ024]

ГТ: Бити на платном листу компаније

РП: Бити на платном списку компаније

У немачком идиому налази се именица Zettel (листић), а у српском именица списак, на основу чега се закључује да постоји лексичка некореспонденција идиома у ова два језика. Како идиом на српском језику гласи бити на платном списку, замена конституената у овом идиому нарушила би критеријум стабилности. Стога се именица листа, коју је програм употребио, означава као недовољно прецизан превод.

На основу анализе грешака увиђа се да је дословни превод најзаступљенија грешка за све типове идиома и да чини 86% свих идентификованих грешака. Може се закључити да је непрепознавање идиома у тексту највећа потешкоћа приликом превода идиома у програму за машинско превођење Google Translate за тестиране примере. У мањој мери заступљене су и грешке настале услед дословног превода идиома са енглеског језика (6%), аномалије у форми преведеног идиома (4,5%), као и непреведени или делимично преведени идиоми (3%).

Прецизни преводи идиома

Као одговор на треће истраживачко питање спроведена је квалитативна анализа прецизних превода. У табели 4 приказане су стратегије превођења у зависности од типа идиома. Стратегије превођења идиома могу бити: превођење идиомом сличне форме и сличног значења, превођење идиомом сличног значења, а различите форме, парафраза и изостављање (Бејкер 1992: 71–78).

Табела 4. Расподела стратегија превођења у корпусу

	Идиом сличне форме и сличног значења	Идиом сличног значења а различите форме	Парафраза	Изостављање
Потпуни идиоми	20	8	20	0
Семи-непрозирни идиоми	13	3	9	0
Семи-транспарентни идиоми	11	0	1	0

У табели 4 може се уочити да је код семи-транспарентних идиома највећи број превода идиомом сличне форме и сличног значења. Анализа прецизних превода потпуних и семи-непрозирних идиома показала је да су и ови идиоми превођени у највећем броју случајева идиомом сличне форме и сличног значења. Са друге стране, код ових типова идиома присутне су и друге стратегије превођења као што су парафраза и превод идиомом сличног значења, а различите форме, на основу чега се може закључити да је Google Translate јасно препознао ове идиоми, као што је то представљено у примерима 10 и 11.

Пример 10.

Seitdem der Herr Bürgermeister ein Nazi geworden ist, geht ihm vieles leichter von der Hand. [TP040]

ГТ: Откако је градоначелник постао нациста, многе ствари су му постале лакше.

РП: Откако је господин градоначелник постао нациста, много му лакше иде у животу.

Пример 11.

Hinter vorgehaltener Hand erzählen [TP051]

ГТ: Причало се иза затворених врата.

РП: Кришом се шапугало.

Квалитативна анализа показала је да су прецизни преводи семи-транспарентних идиома углавном њихови дословни преводи и да преносе у српском на исти начин као и у немачком језику и дословно и идиоматично значење. На основу тога не може се закључити да је Google Translate препознао се-

ми-транспарентне идиоме, већ само да је дословни превод употребљен у преводу већине семи-транспарентних идиома. Овакав прецизан превод семи-транспарентног идиома приказан је у наредном примеру:

Пример 12.

jemanden aus dem Bett holen [ЧЖ074]

ГТ: Дигао га је из кревета.

РП: Извукао га је из кревета.

У примеру 12 приказан је семи-транспарентни идиом који се може интерпретирати и у дословном значењу подићи некога из кревета и у идиоматичном пробудити некога. За оба значења у српском и немачком језику се помоћу овог израза преноси и идиоматично и дословно значење. Превод је у овом случају олакшан због кореспонденције између језика. Тиме је превод идиома олакшан без обзира на способност програма да га препозна у тексту. Стога прецизност у преводу не може бити главни показатељ да је програм препознао овај тип идиома.

На основу квалитативне анализе прецизних превода увиђа се да су код потпуних и семи-непрозирних идиома присутније стратегије превођења које указују на то да је програм Google Translate јасно препознао ове идиоме у тексту. Са друге стране, већина прецизних превода семи-транспарентних идиома представља дословни превод чиме није јасно да ли су ови идиоми препознати у програму. Даљом анализом показано је да постојање лексичке и структурне кореспонденције између српског и немачког језика може допринети прецизности у преводу. Том приликом програм не мора да препозна идиом у тексту, али може да произведе прецизан превод у виду идиома сличне форме и сличног значења.

Закључак

Циљеви овог истраживања били су да се испита прецизност у преводу идиома у зависности од њиховог типа и да се одреде потешкоће које доводе до грешака у преводу са немачког на српски језик у програму Google Translate. У истраживању су идиоми најпре разврстани према типу идиоматичности на потпуне, семи-непрозирне и семи-транспарентне идиоме. Анализа показује да је за тестиране примере програм остварио већи број прецизних превода идиома, приликом чега је једино код семи-транспарентних идиома већи број недовољно прецизних превода. Даља анализа недовољно прецизних превода показала је да је највећи број дословних превода и да ова грешка углавном

настаје услед непознавања идиома у тексту. Ова грешка доминанта је код свих типова идиома. У мањој мери присутне су и грешке попут превода са енглеског језика, приликом чега је идиом испрва идентификован и преведен на енглески језик, а потом дословно преведен са енглеског на српски језик, потом грешка у виду аномалије у форми идиома и непреведен или делимично преведен идиом. Анализа прецизних превода додатно је потврдила да су преводи семи-транспарентних идиома углавном дословни преводи. На основу ових података, може се закључити да готово сви преводи семи-транспарентних идиома у корпусу представљају њихове дословне преводне.

Овакви резултати указују на то да прецизност у преводу идиома није најбољи индикатор успешности програма да препознаје идиоме. Прецизност у преводу може бити резултат постојања лексичке и структурне кореспонденције између изворног и циљног језика, чиме дословни превод представља превод идиомом сличне форме и сличног значења. Приликом испитивања прецизности у преводу идиома треба узети у обзир и поменута ограничења приликом одабира примера за анализу. На овом пољу неопходно је спровести додатна истраживања и том приликом размотрити превод семи-транспарентних идиома који немају преводне еквиваленте у виду идиома сличне форме и сличног значења у циљном језику. Циљани одабир таквих идиома могао би пружити бољи увид у могућност програма да препознаје семи-транспарентне идиоме.

Насупрот томе, потпуни и семи-непрозирни идиоми имају већи број прецизних превода код којих су заступљеније стратегије превођења попут парафраза и превода идиомом сличног значења а различите форме. Ови подаци иду у прилог томе да програм за машинско превођење Google Translate боље препознаје ове идиоме у тексту. Такође, код овог типа идиома учавају се и грешке попут превода са енглеског језика и аномалије у форми идиома у циљном језику које такође показују да је идиом препознат. Ови подаци у складу су са поставкама из теоријског дела (Шторер и Швал 1995; Бејкер 1992), према којима се боље препознају идиоми са већим степеном идиоматичности.

Модификована таксономија за анализу грешака из овог рада може послужити за даљу евалуацију превода идиома у другим програмима за машинско превођење. Такође, ово истраживање, на основу кога се стиче слика о тренутним могућностима и ограничењима програма за машинско превођење Google Translate, може бити подстрек за даља истраживања из компјутерске и примењене лингвистике, која би могла да допринесу побољшању система за машинско превођење у погледу превода идиома.

Литература

Корпус

- Зеталер 2019: R. Zetaler, Trafikant, prev. Lj. Ilić, N. Vukajlović-Fischer, Beograd: Booka.
- Зеталер 2018: R. Zetaler, Čitav život, prev. Lj. Ilić, N. Vukajlović-Fischer, Beograd: Booka.
- Зеталер 2014: R. Seethaler, Ein ganzes Leben. München: Carl Hanser Verlag.
- Зеталер 2012: R. Seethaler, Trafikant. Zürich: Kein & Aber.

Библиографија

- Базотиос и др. 2023: C. Baziotis, P. Marhur, E. Hasler, Automatic Evaluation and Analysis of Idioms in Neural Machine Translation, Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, 3682–3700.
- Бейкер 1992: M. Baker, In Other Words. A coursebook on translation. London: Routledge.
- Буамур 2011: D. Bouamor, N. Semmar, P. Zweigenbaum, Improved Statistical Machine Translation Using MultiWord Expressions. Proceedings of the International Workshop on Using Linguistic Information for Hybrid Machine Translation. 15–20.
- Бургер 2010: H. Burger, Phraseologie. Eine Einführung am Beispiel des Deutschen. Berlin: Erich Schmidt Verlag.
- Коста и др. 2015: A. Costa, W. Ling, T. Luis, R. Corria, L. Coheur, A linguistically motivated taxonomy for Machine Translation error analysis. Machine Translation. 29(2), 127–161.
- Манојловић и др. 2017: M. Manojlović, L. Dajak, M. B. Bakarić, Idioms in state-of-the-art Croatian-English and English-Croatian SMT systems, 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 1546–1550.
- Фернандо и Флавел 1981: C. Fernando, R. Flavell, On idiom: critical views and perspectives. Exeter: University of Exeter.
- Флајшер 1982: W. Fleischer, Phraseologie der deutschen Gegenwartssprache. Leipzig: VEB Bibliographisches Institut.
- Палм 1995: C. Palm, Phraseologie: Eine Einführung. Tübingen: Gunter Narr Verlag.
- Петронијевић 2007: B. Petronijević, Srpsko-nemački prevodni frazeološki rečnik. Beograd: Jasen.

- Прћић 2016: Т. Прчић, *Semantika i pragmatika reči*. Novi Sad: Filozofski fakultet.
- Рамлов и Кренц 2008: М. Ramlow, М. Krenz, *Maschinelle Übersetzung und XML im Übersetzungsprozess*. Berlin: Verlag Frank & amp;
- Салтон 2014: G. Salton, R. Ross, J. Kelleher, *An Empirical Study of the Impact of Idioms on Phrase Based Statistical Machine Translation of English to Brazilian-Portuguese*. *Proceedings of the 3rd Workshop on Hybrid Approaches to Machine Translation (HyTra)*. 36–41.
- Тасић 2023: Љ. Тасић, *Анализа семантичких грешака у програму за превођење Google Translate за језички пар српско-немачки на примеру новинских чланака*, *Зборник радова са научно-стручног скупа Савремена филолошка проучавања младих истраживача I*. Ниш.
- Уану и др. 2018: I. J. Unanue, L. G. Arratibel, E. Z. Borzeshi, M. Piccardil, *English-Basque statistical and neural machine translation*, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*, 880–885.
- Шао и др. 2017: Y. Shao, R. Sennrich, B.L. Webber, F. Fancellu, *Evaluating Machine Translation Performance on Chinese Idioms with a Blacklist Method*, *ArXiv*, abs/1711.07646.
- Шванке 1991: М. Schwanke, *Ein Überblick über Theorie und Praxis*. Heidelberg: Springer.
- Шторер и Швал 1995: А. Storrer, U. Schwall, *Description and Acquisition of Multiword Lexemes*. *Machine Translation and the Lexicon*, 35–50.

Ljiljana Tasić

**THE ANALYSIS OF THE TRANSLATION OF IDIOMS BY
THE GOOGLE TRANSLATE MACHINE TRANSLATION
PROGRAM IN THE CASE OF THE NOVELS ENTITLED
“A WHOLE LIFE” AND “DER TRAFIKANT” FOR THE
GERMAN-SERBIAN LANGUAGE PAIR**

Summary

This paper deals with the analysis of the translation of idioms from German to Serbian by the Google Translate machine translation program. The corpus consists of Robert Zettler’s novels “A Whole Life” and “Der Trafikant”. The objectives of the research were (1) to examine the degree of recognition of idioms depending on their type and (2) to examine the difficulties that lead to errors in the translation of idioms by this program. It was found that Google Translate better recognizes pure-idioms and semi-idioms compared to semi-literal idioms. Error analysis showed that the most prevalent error in the corpus was the literal translation of idioms. Besides this error, errors in literal translation from English, anomalies in form, and untranslated or partially translated idioms were identified. These errors stem from various difficulties, such as failure to recognize idioms in the text, intermediate translation from English to Serbian and lexical and structural discrepancies between German and Serbian. The additional analysis of accurate translations showed that Google Translate uses idiom translation of a similar meaning and similar or different form as well as paraphrases in idiom translation.

Key words: idioms, machine translation, Google Translate, German, Serbian