

TARTU UNIVERSITET
Colleget för främmande språk och kulturer
Avdelningen för skandinavistik

Aleksandra Fomicheva

**GOOGLE TRANSLATE:S ÖVERSÄTTNINGSTEKNIK
FRÅN SVENSKA TILL RYSKA: EN ANALYS AV FYRA
ÖVERSÄTTA TEXTER MED OLIKA BAKGRUNDER**

Bacheloruppsats VT 2015

Handledare: Maiu Elken

Tartu 2015

Soovitan suunata kaitsmisele.....

(juhendaja allkiri)

Kaitsmisele lubatud „ “2015.a.

..... õppetooli juhataja
(õppetooli juhataja nimi ja allkiri)

Olen koostanud töö iseseisvalt. Kõik töö koostamisel kasutatud teiste autorite tööd, põhimõttelised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on viidatud.

.....

(töö autori allkiri)

Innehållsförteckning

FÖRKLARINGAR AV FÖRKOTNINGAR OCH TERMER	4
INLEDNING.....	5
1. TEORETISK BAKGRUND AV MASKINÖVERSÄTTNING OCH GOOGLE TRANSLATE	7
1.1 Maskinöversättning: definition, tekniker och problem	7
1.2 Maskinöversättningens historia av utveckling	9
1.3 Om statistisk maskinöversättning och Google Translate	10
1.4 Skillnader mellan svensk och rysk grammatik	12
1.5 Tidigare forskning.....	14
2. JÄMFÖRANDE ANALYS AV TEXTER ÖVERSÄTTA MED GOOGLE TRANSLATE FRÅN SVENSKA TILL RYSKA.....	17
2.1 Valet av materjal	17
2.2 Metod	19
2.3 Resultat och analys.....	24
2.4 Sammanfattande diskussion	33
LITTERATURFÖRTECKNING	36
BILAGOR.....	40
RESÜMEE.....	51

FÖRKLARINGAR AV FÖRKOTNINGAR OCH TERMER

GT – Google Translate

MT – Maskinöversättning

SMT – Statistisk maskinöversättning

RBMT – Regelbaserad maskinöversättning

BLEU – Bilingual Evaluation Understudy, algoritm för utvärdering av kvalitet av maskinöversatta texter, som jämför outputen med mänsklig översättning

SVO – grundordföljd där subjektet (S) står först i satsen, följt av verbet (V) och objektet (O)

Korpus - inom lingvistik en samling språkliga data som används vid språkforskning, bl.a. statistisk analys

Etalongöversättning / referensöversättning – översättning som skapas av en professionell översättare, ett exempel på korrekt översättning

Direktöversättning – ordagrann översättning utan grundlig analys av olika språkliga nivåer

Transfer – inom maskinöversättning en metod som bearbetar texter på strukturell nivå

Interlingua – inom maskinöversättning en metod där källspråket överförs till en språkneutral representation, som fungerar som en bas för generering av en måltext.

Hybrid-maskinöversättning – en kombination av olika MT-system (t.ex. frasbaserad SMT och exempelbaserad MT)

T#m# - Text (nummer istället för #), mening (nummer istället för #)

INLEDNING

Informationstekniken har börjat att delta aktivt i vårt liv redan för ett par årtionden sedan och nuförtiden upptar det en väldigt stor del av vårt vardagsliv. Vi arbetar, studerar, kommunicerar med hjälp av bl.a. Internet, som underlättar en individs rutin. Emellertid finns det en aspekt som bromsar individens maximala integration inom virtuellt utrymme, som kallas för språkbarriär. Problemet har inte lösts helt för tillfället, inte för vardagslivet, inte heller för cyberrymden. Därför är översättning, som man börjat använda för tusentals år sedan, den enda möjligheten för reducering av språkbrott.

Dock passar inte alltid klassiska metoder nutidens levnadsvillkor, där informationsvolymen ständigt ökar. De traditionella översättningssätten kräver ofta stora finans- och tidstillgångar. I vissa situationer är det enklast att använda maskinöversättning, som gör bearbetningen och översättningen av informationen avgiftsfri och snabb. En av de mest praktiska och lättillgängliga resurserna för maskinöversättning är Google Translate (GT). Det är en gratis webbsida som hjälper att göra översättningar från och till mängder av olika språk i ett ögonblick. Men när det kommer till kvalitén av översättningar gjorda via maskiner och i synnerhet Google Translate, blir man fortfarande tveksam till att bekräfta att sådana översättningar fungerar fullständigt.

I denna uppsats kommer jag att analysera hur maskinöversättning fungerar mellan svenska och ryska inom GT-plattformen. Nämligen kommer det i fokus fyra autentiska svenska texter med olika bakgrunder: ett formellt brev från byggnadsområde, ekonomisk och medicinsk texter, samt en nyhetsartikel och deras översättningar till ryska med hjälp av Google Translate. Anledningen till att använda texter som tillhör olika områden är att sådana texter representerar ett varierande urval av termer, ordförråd och grammatiska konstruktioner, som hjälper att analysera GT möjligheter från olika håll. Det som särskilt intresserar mig är:

- Hur effektivt fungerar översättningen mellan svenska och ryska i GT.
- Vilka slags fel uppträder det inom maskinöversättning mellan svenska och ryska och hur är det möjligt att klassificera eller gruppera dem.
- Hur kunnig är GT på att översätta fackspråk och separata termer från olika vetenskapliga områden.
- Jämförelse av slutresultaten med tidigare forskningar med samma bakgrund.

För att kunna göra jämförande analys, kommer jag att välja bland texter, som redan är översatta av professionella översättare och som ska fungera som etalong innan texterna blir översatta genom GT.

Valet av ämnet grundas på temats aktualitet och relativa nyhet. Fastän man har börjat att använda maskinöversättning redan i mitten av 1900-talet (Trujillo 1999:4), finns det fortfarande inte så många forskningar i detta ämne, i synnerhet när det gäller svenskan och dess samband med andra språk. Problemet med effektivitet av maskinöversättning är grundläggande för språkområdets utveckling i framtiden och just därför är syftet med forskningen att undersöka GT:s möjligheter och kvalité samt identifiera vilka slags texter GT översätter mest noggrant.

Uppsatsen kommer att bestå av två delar. Den första delen ska handla om teoretiska aspekter, som gäller bl.a. bakgrunden för maskinöversättning och GT, förklaringar om olikheter mellan svensk och rysk grammatik. Samt kommer jag att återvända till tidigare forskningar och sammanfatta resultat, som man har fått förut. Den andra delen av uppsatsen kommer att behandla valet av material och beskriva metoder som har använts för att uppnå syftet. Texter översatta i GT ska analyseras och jämföras. Resultatet kommer att beskrivas och sammanfattas.

Jag vill tacka först och främst min handledare Maiu Elken för värdefulla råd och påpekande. Dessutom vill jag tacka översättaren Olga Alferova vem som hjälpte med referensöversättningarna för alla fyra texter.

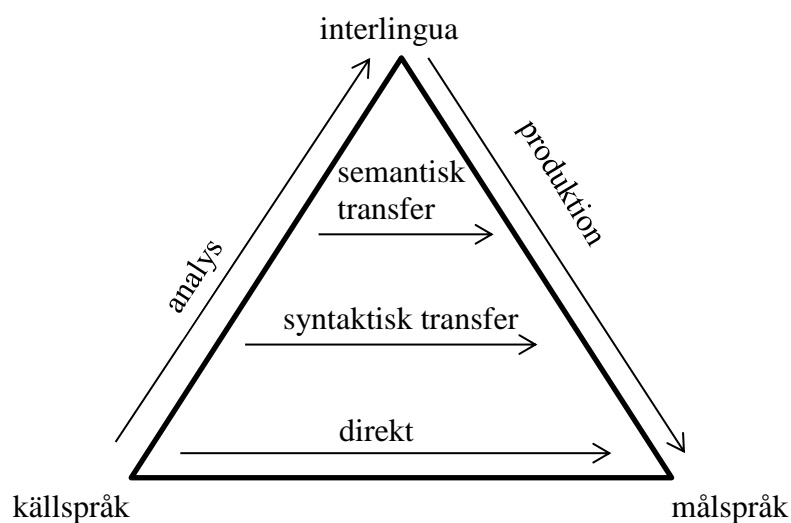
1. TEORETISK BAKGRUND AV MASKINÖVERSÄTTNING OCH GOOGLE TRANSLATE

1.1 Maskinöversättning: definition, tekniker och problem

Maskinöversättning (MT vidare) gäller användning av datamaskiner för att erbjuda en högvärdig och idiomatisk översättning mellan olika par av naturliga språk (Bird 2009:29). Trujillo beskriver MT i sin bok från 1999 som ett hjälpmedel för översättning via olika datamaskiner. Författaren tillägger att ursprungliga texter, som man vanligen använder på MT-plattformen är av icke-litterära typer, där det viktigaste kriteriet för översättning är att behålla innehållet av en text. (Trujillo 1999: 4) På 2000-talet fick maskinöversättning större popularitet bland vanliga användare genom vissa service (Babel Fish, Google Language Tools, Windows Live Translator) som är i regel tillgängliga genom söktjänster (Goutte 2009: 1). Användning av MT är alltså inte något extraordinärt nuförtiden, men man utnyttjar dess möjligheter vardagligt för att översätta texter av olika slag, från litterära textavsnitt till tekniska manualer.

Nuförtiden representeras MT genom olika modeller: något enklare modeller som *direkt översättningsmetod* som konverterar texten till målspråket enligt basreglerna, mer avancerade *transfermetoder* som inefattar redan morfologisk och syntaktisk analys och *interlinguametoder* som är av högsta nivå och erbjuder tolkning av abstrakta betydelser (Koehn 2010: 15). Maskinöversättningens tekniker är representerade i Vauquois pyramid eller den så kallade klassiska MT pyramiden (Figur 1).

Traditionell MT gäller konvertering av en text skriven på ett källspråk till en ekvivalent text på ett annat språk, ett målspråk. Traditionell MT är baserad på olika nivåer av språklig analys på källspråkets sida och språklig produktion på målspråkets sida. (Goutte 2009: 2) Enligt pyramiden i Figur 1 är det klart att transfereffekten är minimal på apex medan analys och produktion är maximala.



Figur 1. Maskinöversättningens pyramid (Trujillo 1999: 6)

Alltså saknar direkta översättningar en fullständig analys, medan texter, som är översatta via interlingua bearbetas med en full analys. Därför att interlingua är en universell representation av alla naturliga språk behöver den ingen transferfas för textkonvertering och således ligger interlingua på toppen av pyramiden. Trujillo påpekar också att pyramiden är idealiserad och att olika MT-verktyg i praktiken använder olika variationer av strategier. Det kan vara t.ex. kombinationer av direkt översättning och transferelement som är en del av statistisk maskinöversättning och andra korpus-baserade tekniker. (Trujillo 1999: 6).

Men inte bara tekniker återspeglar maskinöversättningens struktur. Man ska också tänka på informationskällor som MT använder. I sin forskning förklarar Costa-jussà och Fonollosa att det ingår *regler* i informationskällor, som är en lingvistiskt motiverad faktor, och *data*, som är baserad på statistik. MT som är baserad på regler använder lingvistisk information (t.ex. monolingvistiska och bilingvistiska ordböcker) i samband med mänskliga lingvistiska kunskaper. Data-baserade MT använder informationen från data och komplexa algoritmer vilket resulterar i modell-baserad översättning. Sådana MT är i regel exempelbaserade (EBMT) eller statistiskt-baserade (SMT). EBMT översätter en text direkt i analogi med tidigare schabloner vilket ibland kan vara problematiskt, därför att schabloner inte passar vid varje fall. (Costa-jussà, Fonollosa 2015: 4) SMT beskrivs helt annorlunda i samma forskning: “SMT systems try to find the most probable translation given the source sentence, by reference to the models built

using data such as the translation and language model.” (*Ibid.*: 4) Om man summerar det ovannämnda, då är det de olika teknikerna som är viktigaste inom MT (direkta översättningar, transfer och interlinguametoder) och informationskällorna (som representeras av regler och data).

Ett stort subjekt för diskussionen har alltid varit MT:s nyttighet och funktionalitet. Det finns huvudsakligen två fördelar hos maskinöversättning – den brukar vara både snabbare och billigare än manuell översättning. Nämligen finns det flera gratis webbsidor, som är baserade på MT-plattformen och som kan översätta texter med ett enda klick. Det kan spara tid när man behöver att översätta stora informationsvolymer, som exempelvis en webbsidas innehåll eller en förfrågan i ett söksystem. Problemet som påträffas oftare vid maskinöversättning är att det ibland inte motsvarar sin definition inkl. „att erbjuda en högvärdig och idiomatisk översättning“. Maskiner har det svårt att konvertera en text till ett målspråk med alla språkliga nyanser, det förekommer olika fel vid översättningen och slutresultatet kan vara konstigt. Hutchins bekräftar MTs brister, när han säger att om MT används för att göra översättningar av högre kvalitet t.ex. för publicering (liknande med mänskliga översättningar) måste slutresultat redigeras, eller som sägs *postedited* (Hutchins 2003: 502). Alltså är mänsklig inblandning oundviklig i detta fall och då skiljer sig MT inte så mycket från traditionella översättningar. Hutchins tillägger också att kvalitét av MT kan förbättras när källtexter ska anpassas genom bl. a. *pre-editing*, som hjälper att identifiera prefix-, suffix-, fras- och satsgränser, skiljetecken samt grammatiska kategorier. (*Ibid.*: 502-503)

1.2 Maskinöversättningens historia av utveckling

Fastän MT:s historia inte är så lång är den ganska intensiv tack vare konstant utveckling i IT-området. Trujillo påpekar att några idéer, som påverkade utvecklingen av MT upptäcktes redan innan datorns uppkomst. Åtminstone redan på sextonhundratalet erbjöd man att använda „language-neutral representations of meaning in order to overcome linguistic barriers“ (Trujillo 1999: 4). Mycket senare, i 1933 uppfanns det första mekaniska verktyget för att producera översättningar av den sovjetiske forskaren Petr Smirnov-Trojanskii. (Trujillo 1999: 4) De viktigaste aspekterna av MT-utveckling representeras vidare i Tabell 1.

Tabell 1. Viktiga aspekter av utveckling av MT under olika perioder

Period	Viktiga aspekter av utveckling av MT
1940-talet	Warren Weaver initierade verksamheten för utveckling av MT när han erbjöd specifika sätt att översätta texter via datorer i sin promemoria (Hutchins 2000: 17).
1950-talet	Intresset för MT var väldigt stort. År 1954 startades det stora IBM projektet som översatte 49 meningar från ryska till engelska, fast systemet fungerade dåligt och översatte meningar direkt (Hutchins 2010: 30).
ALPAC- era	Automatic Language Processing Advisory Committee (ALPAC) inställde sponsring av olika MT-projekt p.g.a. dåligt kvalitet och irrationalitet (Hutchins 2010: 33-34).
1970-talet	Systran började översätta texter från engelska till franska Samtidigt översatte Météo-systemet väderleksprognoser. (Trujillo 1999: 5)
1980-talet	Logos och METAL system skapades för allmän tillämpning, fastän i praktiken anpassades deras ordförråd till vissa fack (Koehn 2010: 16). Dessutom kunde man anpassa en trestegs transfer-baserade approach med orientering på syntax och förbättring av lexikala och grammatiska regler. Det mest kända projektet från 1980-talet var Eurotra-projektet av Europeiska gemenskaperna. (Hutchins 2010: 36)
1990-talet	Exempelbaserade översättningssystem infördes, som försökte att hitta en mening som liknade meningen i en parallell korpus och anpassade förändringar som behövdes till en redan existerande översättning. Det var en bas för SMT. (Koehn 2010: 17)
Internet- era	I mitten av 1990-talet påverkades maskinöversättningens inriktning mycket av utvecklingen av Internet . Först fanns det några verktyg (Systran, BabelFish) som översatte webbsidor och elektroniska meddelanden. Verkygen var populära fast kvalitét av on-line MT-service var mycket låg . På 2000-talet blev SMT väldigt aktuell (Hutchins 2010: 45).

Nuförtiden försöker traditionella MT företag som Systran att integrera statistiska MT-modeller med deras system. Det visar också nyttighet och aktualitet av MT, därför att Internetanvändare översätter 50 million webbsidor per dag via system som Google, Yahoo, Microsoft osv. (Koehn 2010: 18) Alltså är MT-system på övergångsperiod till statistiskt och korpus-baserade hybrider, som är av högre kvalitét, vilket ger oss hopp att maskinöversättningar ska fungera perfekt i framtiden.

1.3 Om statistisk maskinöversättning och Google Translate

Statistisk MT blev populär på slutet av 1990-talet, den var en upptäckt som skulle förnya och förändra attityd till MT och införa helt annorlunda approach. Hu påpekar att

statistisk maskinöversättning (SMT) i viss mån försöker att hitta översättningen genom statistisk beräkning, som förstärker regler och exempel SMT jobbar på och gör dem mer sannolika (Hu 2014: 1628). SMT skiljer sig från den traditionella, regel-baserade MT därför att det fungerar enligt statistiska modeller och algoritmer, fast båda använder för översättning olika fraser uppladdade i parallella korpusar. Menezes och Quirk skriver att SMT:s system är bra på att producera korrekta och idiomatiska översättningar på lokal nivå, fast det har problem med flera lingvistiska företeelser: exempelvis presenteras grammatiska företeelser ganska dåligt (Menezes, Quirk 2005: 99). Dessutom finns det problem med stil i SMT, vilket beror på databas. Syahrina beskriver den företeelsen som ”domain flexibility” och förklarar att:

The problem with domain flexibility is that different domain might have different style of language. A machine translator that does not recognize domain flexibility might recognise the sentence as a mistake, or translate wrongly. For example, there would be a difference in translating formal language and informal language. <...> However for statistical machine translation, it needs to be trained with new set of data in a different domain. (Syahrina 2011: 23)

Det är anmärkningsvärt, hur SMT kombineras med alla andra MT-system på sista tiden och idag använder man oftare begreppet *hybrid maskinöversättning*. När det gäller korpusbaserade system, så är det exempelvis frasbaserad SMT och exempelbaserad MT som kombineras (Costa-jussà, Fonollosa 2015: 6). Sådana kombinationer hjälper att förbättra MT, därför att hybrider syftar på användning av olika MT-system samtidigt. En av de MT-verktyg, som försöker att anpassa sitt system till olika MT-tekniker och strävar efter bättre kvalitet är Google Translate.

Google Translate är en tjänst som hjälper att översätta texter eller webbsidor till ett annat språk. GT erbjöds av en stor söktjänst Google Inc, det omfattar 90 olika språk, inklusive planspråket Esperanto och klassisk Latin och det fungerar sedan år 2001 (Google Company 2015). När GT började sin verksamhet, baserades verktyget på Systrans regel-baserade modeller som alla andra MT-system använde, men Google Inc bestämde att förändra sin MT-modell under tidens lopp. Enligt en artikel på Google Operation System:s webbsida övergick GT från Systran:s modell till sitt eget MT-system med statistisk maskinöversättningsapproach år 2007 (Google Switches... 2015). Barreiro et al. påpekar att GT är mest populär online MT-system bland Internet-

användare idag och att det grundas på SMT-system, som har mängder av parallella data, som Google samlar ifrån webben. Det är typiskt för GT – ett system som baseras på SMT – att skapa bättre resultat vid översättning mellan nära språkpar (t.ex. portugisiska och spanska) och mellan språk som har mycket information laddade upp i parallella korpusar. Alltså är det mycket beroende av språkpar. Å andra sidan är GT ett slutet system och det är oklart om det finns någon semantisk förståelse inom GT. (Barreiro et al. 2014: 35-36) Samtidigt påpekar Groves & Mundt att personlig uppfattning av GT:s kvalité beror på syften: om man använder GT för att läsa någonting på ett främmande språk, accepterar man översättningar med några fel, därför att man begiper vad texten handlar om generellt. Emellertid om man förväntar sig att programmet skapar en text för en läsare av ett målspråk, finns det i regel högre behov av grammatisk noggrannhet. (Groves & Mundt 2015: 115)

På Google Translate:s egen webbsida kan man följa uppdateringar av tjänsten. Idag erbjuder Google Inc att ladda upp GT som app till mobiltelefon: det är inte nödvändigt att skriva texten till *input*, man kan använda sin röst och även ta bilder av inskrifter på ett främmande språk – då får man översättningen direkt i sin telefon. Dessutom praktiserar GT interaktivt samarbete med sina användare: man kan göra korrigeringar direkt i produkten om man ser en felaktig översättning och således förbättras innehållet av GT korpusar. (Google Translate 2015) Google Translate är en av de tjänster, som försöker att anpassa sig till framstegets kriterier: det gäller inte bara tekniska korrigeringar nämnda ovanför, men också en kontinuerligt förbättring av data i korpusar vilket resulterar i bättre översättningar. Nuförtiden är GT:s översättningar inte ideala, fast mycket bättre än för till och med 2-3 år sedan. Just därför är det intressant och nyttigt att följa utveckling av GT och jämföra det med tidigare forskningar.

1.4 Skillnader mellan svensk och rysk grammatik

Både svenska och ryska tillhör de indoeuropeiska språken – den största språkfamiljen i världen. Ryska klassificeras som det största av slaviska språken och det är åttonde språket i världen med 166 miljoner antal talare (Ethnologue 2015). Inom ryskan används det kyrilliska alfabetet. Svenskan, å andra sidan, är ett germanskt språk, som

klassificeras som nordiskt och har ca 10,5 miljoner antal talare (Ethnologue 2015). Svenskan skrivs med en variant av det latinska alfabetet.

Fastän ryskan och svenskan har några liknande drag i sina grammatiska strukturer, är större delen av dem olika. Andersson skriver att det är kasus som spelar en stor roll (Andersson, L. 1992: 55). Svenskan har två kasus för substantiv: grundkasus och genitiv (t.ex. *pojke; pojkes*). För personliga pronomen finns det också en tredje kasus, ackusativ (t.ex. han – hans – honom; hon – hennes – henne). (Andersson, E. 1994: 28; 68) Ryskan har däremot sex kasus: nominativ, genitiv, dativ, ackusativ, instrumentalis och lokativ. De påverkar inte bara substantiv, men också adjektiv och pronomen.

Ryskan och svenskan har det gemensamt att adjektivattributet i regel står framför sitt huvudord, sällan efterställt (Andersson, E. 1994: 90; 93). Dessutom har båda språken i princip samma ordföljdsregler, båda tillhör s.k.. SVO-språk (subjekt – verb – objekt) (*Ibid.*: 142-143). Men till skillnad från svenskan bestäms ordföljden i ryskan inte lika högt. Olika ordföljder används i ryskan t.ex. för att betona vad som är viktigt i sammanhanget eller för att betona att någonting är nytt och okänt eller tvärtom gammalt och känt. I svenskan uttrycks detta genom bestämd eller obestämd form (t.ex. *Jag såg en häst; Jag såg hästen*). I ryskan saknas bestämd och obestämd form på substantiv, som uttrycks i svenskan genom bl.a. artiklar. (Andersson, L. 1992: 56) Svenska språket har alltså två genusgrupper: utrum (*en bok*) och neutrum (*ett ord*), beroende på vilken bestämdhetsändelse de tar i singularis (Andersson, E. 1994: 30). Som alla andra slaviska språk har ryskan tre genus: maskulinum – femininum – neutrum. En av de viktiga skillnaderna mellan ryskan och svenskan är placering av genitivattribut. I svenskan står genitivattributet alltid före huvudordet (fetstil). Däremot står genitivattributet i ryskan oftast efter huvudordet (Andersson 2001: 17), t.ex. *kattens ögon – глаза кошки; Marias make – муж Марии; Sveriges huvudstad – столица Швеции*.

Olsson påpekar att negationsord står på olika ställen i svenskan och ryskan. I ryskan placeras det alltid före huvud verbet (fetstil): t.ex. *Я не хочу спать* (direkt översättning: *Jag inte vill sova*). I svenskan står negationsorden däremot på sin fasta plats, dvs. efter det finita verbet i huvudsatser och före det finita verbet i bisatser. Dessutom är det ganska vanligt för ryskan att utelämna kopulaverbet (motsvarar svenskans *är*) eller subjektet, som är obligatoriska i en svensk mening. Om kopulaverbet utelämnas i

svenskan, kan satsen inte fungera på ett grammatiskt sätt, medan utelämning av subjektet leder till svårigheter i att förså meningens meddelande. (Olsson 2013: 10)

Det finns huvudsakligen ganska många grammatiska skillnader mellan svenska och ryska, trots att båda tillhör indoeuropeiska språk. Det är inte alltid lätt att anpassa språkliga nyanser och grammatisk varians för en mänsklig översättare, så det är dubbelt intressant att iaktta hur det fungerar i ett maskinöversättningssystem.

1.5 Tidigare forskning

Det finns inte så många forskningar angående maskinell översättning mellan svenska och ryska, speciellt när temat är begränsat med en viss sorts MT – Google Translate. Därför valde jag att behandla tidigare forskning från en annan synpunkt: hur översättningar i GT fungerar mellan olika språkpar. Jag försökte att koncentrera sig på stickprovet som innehåller åtminstone ett språk, som behandlas i denna uppsats (alltså ryska eller svenska) eller som pekar på ett specifikt område (i översikten är det medicin). Kortfattade översikt om metoder framläggs i Tabell 2.

Tabell 2. Kortfattade översikten av andra undersökningar om översättning i GT

Språkpar	Författaren och år	Metod
Ryska-svenska	Aleman Tennell, 2013	Analys av 14 korta textstycken och 19 exempelmeningar, för att se hur GT hanterar rysk ordföljd.
Svenska-estniska	Hiire, 2013	Komparativ analys, jämförande fel analys av tre olika texter översatta via GT
Svenska-danska	Francis & Nordfalk, 2009	Komparativ analys mellan tre MT system, inkl GT.
Engelska-svenska- engelska	Syahrina, 2011	Jämförande analys av två MT system: GT och Systran på olika textavsnitter.
	Salimi, 2014	BLEU analys av litterära (10 stycken) och icke-litterära texter (10 stycken) översatta via GT.
	Allansson, 2014	Analys av översättningen av svenska sammansättningar via GT i nio texter.
	Askarieh, 2014	Fel analys av 2 texter översatta via GT och Bing.
Engelska-... 13 europeiska, 11 asiatiska och 2 afrikanska språk	Patil & Davies, 2014	Tio medicinska fraser utvärderades i 26 olika språk, vilket resulterade i 260 översatta fraser, resultaten jämfördes med mänsklig översättning.

Sammanfattande resultat om tidigare forskning kan man hitta i Bilaga 5. Det är lättare att följa forskningens resultat, om det finns ett konkret värde som uttrycks procentuellt. Då har man en tydlig bild om GT:s funktionalitet. I några forskningar beskrivs resultatet inte så exakt, vilket gör värderingen av GT mer subjektiv. Jag kunde inte hitta undersökningen, där resultatet av texter översatta genom GT skulle vara perfekt (100 procent begripligt eller nära det), fast dåliga resultat, då översättningar är nästan obegripliga, ibland förekommer. Den enda undersökningen som gäller språkparet svenska-ryska är *Översättning från ryska till svenska i Google Translate* av Maximilian Aleman Tennell, 2013. Den visade att GT ofta väljer att direktöversätta ryskans ordföljd till svenska och om ryskans ordföljd avviker från SVO-mönstret blir översättningarna inkonsekventa. GT upfattar inte ryskans kasusändelser och blandar samman placering av negationsordet *inte* samt genitivattribut. Några av samma problem uppträder också i ett svenskt-engelskt språkpar, som beskrivs i undersökningen *Cohesion and Comprehensibility in Swedish-English Machine Translated Texts* av Sona Askarieh, 2014. Författaren hävdar, att vid översättningen från svenska till engelska har det funnits små fel som har gällt prepositioner och placering av genitivattribut. Undersökningen som gäller medicinskt område (*Use of Google Translate in medical communication: evaluation of accuracy* av Patil & Davies, 2014) har visat att GT inte passar så bra för viktiga och legala medicinska samtal, därför att hos vissa språkpar (speciellt sådana som innehåller afrikanska och asiatiska språk) är översättningen helt inkorrekt och kan tolkas fel, vilket kan krångla till kommunikationen mellan läkare och patient.

Sammanfattningsvis kan man dra slutsatsen, att även om det presenteras olika språkpar i tabellen, skiljer sig resultaten inte så mycket åt: för det mesta är googleöversättningarna begripliga, speciellt om de är av icke-litterära slag, men GT har svårigheter med några grammatiska konstruktioner och enskilda ord i nästan varje översättning. Nästan varje författare tillägger också, att även om det finns fel vid GT-översättningar, är det fortfarande ett nyttigt och snabbt vis för översättning och jämfört med andra MT-system placeras Google Translate på den högsta nivån.

Den första delen av uppsatsen handlade om teoretisk bakgrund för maskinöversättning och Google Translate. Jag behandlade definitionen av MT, problem som förekommer,

främsta tekniker och informationskällor. De viktigaste modellerna av MT observeras genom historisk utveckling av MT. Statistisk maskinöversättning är huvudbasen för Google Translate och dennas princip beskrivs i kapitel 1.3. I samma kapitel får man veta om Google Translate som online översättningstjänst. Det var också viktigt att peka på de största skillnaderna i grammatiken mellan svenskan och ryskan, för att förstå naturen av fel och begränsningar, som kan förekomma vid översättning med GT. Slutligen betraktade jag tidigare forskning om GT, vilket ger en översikt av problem och fel som uppträder vid användning av GT samt hjälper att förstå den ungefärliga nivån av GT:s funktionalitet. Dessa forskningar kan jämföras med resultat som jag får genom analys av olika texter översatta med GT i nästa del av uppsatsen.

2. JÄMFÖRANDE ANALYS AV TEXTER ÖVERSATTA MED GOOGLE TRANSLATE FRÅN SVENSKA TILL RYSKA

2.1 Valet av materjal

Med denna undersökning kommer jag att klarlägga hur effektivt översättningen mellan svenska och ryska med GT fungerar och vilka fel det förekommer i översättningens process. För mig är också viktigt att kontrollera hur kunnig GT är på att översätta fackspråk och separata termer från olika vetenskapliga områden. Således bestämmer syftet med undersökningen valet av materjal för analysen. Det blir sammanlagt fyra svenska texter med olika bakgrunder som ska användas för översättningen med GT.

Textlängderna varierar från 206 till 219 ord, så det är viktigt att påpeka att stickprovet är relativt små för att återspegla hela verksamheten av GT och omfatta all information som finns i GT:s korpusar. Resultaten ska bara tolkas som bekräftelse av vad som kan vara typisk för GT. Texternas kvalitet spelar också en ganska viktig roll: alla fyra texter har olika teman och bakgrunder, några av dem är mer formella än de andra. Några textavsnitt har valts därför att de representerar ett omfattande urval av specifika termer och fackspråk, så att det skulle bli möjligt att testa Google Translate:s kunnighet att översätta detta. En av texterna är av enklare journalistiska slag, som skapar kontrasten till textavsnitten med vetenskaplig eller specifik bakgrund, vilka är svårare att uppfata för mottagaren utan kunskaper i ett konkret område.

Den första texten är ett avsnitt av ett semi-formellt brev, som handlar om redovisning och bedömningsunderlag för en speciell produkt från byggnadsområde. Textens källa är Olga Alferovas privata samling av översatta texter. Texten är ganska specifik därför att det uppträder många termer från byggnadsområdet (t.ex. *tvärsnitt*, *lindningsavstånd*, *konstruktionsberäkningar*, *drifts-och skötselavvisningar* osv.), som inte är lätta att förstå

för en vanlig läsare. Texten består av 206 ord och efter översättning med Google Translate till ryska har det blivit det 200 ord.

Den andra texten är ett avsnitt ur en artikel om tuberkulos på Wikipedia (Tuberkulos 2015), som innehåller 219 ord och den ryska GT versionen av texten innehåller 208 ord. Ämnesområdet är alltså medicin och därför innehåller texten många medicinska och biologiska termer (som t.ex. *tuberkelbakterien*, *makrofager*, *centrala nervsystemet*, *tuberkulintest på huden* osv.). Det finns en förkortning i texten – *BCG-vaccin* – som kanske kan vara problematisk för GT att hantera. Å andra sidan är den syntaktiska delen inte så komplicerad. Texten består huvudsakligen av enkla och korta meningar. Det finns meningar med bisatser eller två huvudsatser som ändå har en ganska enkel konstruktion. Detta underlättar översättningsarbetet för GT.

Den tredje texten har en ekonomisk bakgrund och handlar om inflation och dennas verkningar (Inflation 2015). Textavsnittet kommer också från en artikel på Wikipedia, det innehåller 207 ord och dess översättning till ryska innehåller 202 ord. Texten har flera ekonomiska termer (t.ex. *penningmängden*, *konsumentprisindex*, *åtstrammingspolitik* osv.) men dess syntaktiska konstruktion är också okomplicerad. Det finns bara ett par meningar, där det kan vara ganska svårt för GT att hantera översättningen på grund av meningarnas längder och flera bisatser i samma mening.

Den fjärde och sista texten gäller en texttyp som brukar vara lättare för uppfattning av en vanlig läsare – det är en artikel på SvD:s nyhetsportal om internetköp och människors missnöje med det (Så slipper du 2015). I artikeln används det å ena sidan ett vardagligt, enkelt språk (exempelvis utelämnning av hjälpverbet *har* i verbets perfektform - *den som köpt något*) med direkt tal. Å andra sidan kan det försvåra översättningen för Google Translate om det påträffas vardagliga ord och talspråkets detaljer, som GT inte kan identifiera och jämföra med informationen i sina korpusar. Texten består av 216 ord som blev 211 efter översättningen till ryska med GT.

För att kontrollera kvalitét av översättningar i Google Translate var alla fyra texter tidigare översatta av en mänsklig översättare, så det ska bli möjligt att jämföra googleöversättningarna med etalongöversättningarna och uppenbara brister i texterna översatta av GT med större precision. Det underlättar också analysen, därför att det ska

bli möjligt att undvika tveksamheter vid googleöversättningar och bestämma om GT har valt en korrekt betydelse för en term, om det redan finns en konkret konstruktion översatt av professionalen. Å andra sidan kan det förekomma flera korrekta översättningar för en källmening (speciellt i ryskan, där det finns många synonymer och möjligheter att variera ordföljden), vilket kan vara problematiskt när man utvärderar korrekthet av maskinell översättning. Om en googleöversättning avviker från referensöversättningen måste det kontrolleras noggrant om tolkningen erbjuden av GT passar kontexten. Ifall de passar så ska den inte räknas som felaktig. Förutom etalongöversättningen används det ytterligare en grammatisk informationsportal, GRAMOTA.RU, som har en stor databas av olika ryska ordböcker och grammatiska regler.

2.2 Metod

Det finns flera metoder och även separata program (t.ex. BLEU) för utvärdering av maskinöversättningens output, inom vilka det används matematiska algoritmer och statistiska prov. Sådana program är bra för jämförelse av olika MT-system. Mestadels erbjuder sådana program ett allmänt resultat och således är det svårt att förstå och tolka varifrån de flesta felen kommer. För att förstå vilka huvudproblem och fel det förekommer behövs det att genomföra en analys av outputen av maskinell översättning. I denna undersökning ska jag inte heller använda något statistiskt program, för stickprovet av källtexter är ganska litet och behöver inte bearbetning med speciella program. Därtill är undersökningens syfte att utreda vilka sorts fel det förekommer och möjligtvis att gruppera dessa fel, vilket framför allt kräver manuell behandling.

Google Translate:s webbsida används som primarkälla för analysen. Webbsidan har två rutor: den första är för texten på ett källspråk eller s.k. *input* och den andra är för översättningen som GT producerar på ett målspråk, dvs. *output*. Till input kopieras det svenska texter som översätts till ryska med GT:s verktyg. GT väljer själv bland alternativen av översättningen, som enligt dess inre korpusar och mekanismer passar bäst. Detta betyder att översättningen genomförs utan mänsklig inblandning eller en s.k. *postediting*. Resultaten sammanställs till tabeller, tillsammans med originaltexterna på svenska och mänskliga översättningar. För att lättare hitta nödvändiga grammatiska

exempel i materialet är varje mening i de svenska texterna numrerad och dess referensöversättning på ryska står bredvid den i nästa spalt. I den tredje spalten står samma mening översatt av GT. Tabellerna med texterna finns i Bilagor 1 – 4.

Googleöversättningarna ska läsas och jämföras med referensöversättningarna och ifall GT:s output avviker från den mänskliga översättningen ska avvikelserna kontrolleras, och om det blir en felaktig grammatisk konstruktion eller ett inkorrekt ordval, ska det markeras som fel och placeras till en av felkategorierna. Felen är grupperade i flera kategorier, som baseras på artiklarna av Daems et al. (2014), Daems & Macken (2013), och Vilar et al (2006). Kategoriernas benämningar står på engelska. Det görs med tanke på framtida forskningar då det blir lättare att jämföra undersökningar och hitta konkreta resultat.

Primarkällan för valet av kategorierna var instruktionen för utvärdering av maskinöversättningens fel mellan engelska och nederländska av Daems & Macken (2013). De erbjuder att använda sex huvudkategorier, som har sina olika underkategorier. En översikt över kategorierna och underkategorierna som jag kommer att använda i analysen finns i tabell 3.

Tabell 3. Översikt av felkategorier och dess underkategorier

Category	Subcategory	Category	Subcategory
Grammar & syntax	incorrect case	Spelling, typos and punctuation	capitalization
	word order		punctuation
	verb form	Style and register	disfluent sentence/construction
	subject-verb agreement		register
	other incorrect endings		repetition
	missing constituent or preposition		a slight inaccuracy
Lexicon	incorrect word	Coherence	missing info
	missing word		the meaning is lost for construction/sentence
	wrong preposition	Adequacy	terminology
	wrong word class		
	extra word		

I artikeln av Daems et. al påpekas det att deras kategorisering är lagom flexibel och passar också för tilläggning av olika underkategorier beroende av syftet med analys och

språkparet. I ryskan är det kasussystemet som är viktigt och som skiljer sig mycket från det svenska kasussystemet, vilket gör att det kan införas underkategorin ”incorrect case” till kategorin ”grammar & syntax” (Daems et al. 2014: 62). Samtidigt kan det utelämnas flera underkategorier från instruktionen, därför att de är anpassade till nederländska språket. Exempelvis passar underkategorin ”article” inte ihop med språkparet svenska-ryska, därför att det inte finns några artiklar i ryskan. Det är viktigt för analysen att särskilt kontrollera översättningen av termer och därför finns det en separat underkategori ”terminology”. Det visade sig dock under analysen, att det är nödvändigt att komplettera kategorierna med några nya underkategorier. Således kommer några extra underkategorier, som gäller främst lexikologi (”incorrect word”, ”missing word”, ”extra word”), från artikeln av Vilar et al. (2006: 699). Några andra underkategorier infördes av mig när det fanns ett behov av att beskriva en språklig företeelse, men det fanns ingen konkret underkategori i källorna. Sådana underkategorier är *other incorrect endings; wrong word class; a slight inaccuracy; the meaning is lost for construction/sentence*. Beskrivningarna av varje kategori och dess underkategorier kan man hitta nedan.

Grammar & syntax betecknar företeelser som inte motsvarar grammatiska och syntaktiska regler av det ryska språket. Under analysen infördes det sex underkategorier till *Grammar & syntax* som bildar den största gruppen av fel. *Incorrect case* betecknar problem med kasus. Exempelvis är det „инструкция по хранению“ (sv. *anvisningar för lagring*) i GT versionen (T1m3.6). Ordet *хранению* böjs i lokativ kasus enligt GT, fast det måste böjas i dativ. *Word order* pekar på inkonsekvent ordföljd i satsen. Fastän det inte är viktigt för ryskan att ha en bestämd ordföljd, finns det några konstruktioner i googleversionerna, där ordföljden krånglar uppfattningen av texten. Det måste också påpekas, att ifall ordföljden är korrekt, men en annan ordföljd skulle passa bättre till stilen eller kontexten, markeras det som *disfluent sentence/construction* fel i kategorin ”style and register”. *Verb form* betecknar inkorrekt tempus eller form av verb, som i t.ex. T4m11 där verbet ”укреплять” måste vara ”укрепить” därför att det skall beteckna en avslutad handling, inte processen. *Subject-verb agreement* visar avvikelser mellan numerus i subjekt och verb, exempelvis T1m5 då subjektet ”признание” står i singularis och predikatet (verbet) ”должны быть доступны” står i pluralis. *Other incorrect endings* används för alla andra felaktiga ändelser, t.ex i T3m2:

”Оригинальный определене” där det finns missanpassning mellan adjektivets och substantivets genus som uttrycks genom ändelser. Underkategorin *Missing constituent or preposition* betyder att meningen kan vara grammatiskt fel därför att en nödvändig del av en konstruktion saknas – det kan vara en preposition eller några längre komponenter som, exempelvis, sammansatta konjunktioner. Exemplet är från T2m2: ”<...> но также может атаковать другие органы, # центральную нервную систему <...>” där tecken # visar att det saknas konjunktionen ”такие как” som krävs när uppräkningsen följer.

Lexicon är den andra kategorin som beskriver lexikala problem eller fel. Feltypen *Incorrect word* markeras när GT inte kan hitta en korrekt översättning av ett konkret ord i källmeningen. I T1m5 är det exempelvis „redovisning“ som översätts fel av GT som „признание“, trots att den korrekta översättningen är „отчёт“. I artikeln av Vilar et al. (2006) kan underkategorin *Incorrect word* delas i flera andra underkategorier men det görs inte i denna analys, därför att det är onödigt. Fel markeras som *Missing word* när ett ord i den översatta meningen saknas. Underkategorin *Wrong preposition* används när en preposition inte passar uttrycket och måste ersättas med en annan preposition eller när prepositionen inte alls behövs. I T4m13 måste prepositionen „на“ ersättas med „в“. *Incorrect word class* är en av de kategorier som har införts av mig och den visar att det är fel ordklass som används för översättning av ett ord. I T2m12 kan man observera att ordet „cellen“ som är ett substantiv i svenskan har översatts som ett adjektiv till ryskan - „клеточных“, vilket är omotiverat och fel.

Spelling, typos and punctuation är en kategori som visar att någonting inte motsvarar regler av stavning eller kommatering i ryskan, eller att det finns ett tryckfel. Emellertid måste det påpekas att det saknas underkategorin *typos* (tryckfel), därför att de ursprungliga texterna är grammatiskt korrekta och felfria, bl.a. finns det inga tryckfel, vilket utesluter möjligheten att tryckfelen kan vara överförda till måltexterna. Kategorin har två underkategorier. *Capitalization* visar att ryskans versaliseringsregler följs inte, som i T1m7 där „Тестирование“ måste stavas med en liten bokstav. *Punctuation* pekar på att punktueringen i meningarna är fel eller saknas.

Style and register. Den kategorin betecknar problem med stilistiska företeelser och ordval. Med underkategorin *disfluent sentence/construction* menas det hos Daems &

Macken (2013) att 'the sentence/constituent is not grammatically incorrect, but it is nonetheless very difficult to read, it could be translated in a much more idiomatic way.' (Daems & Macken 2013: 17). Exempelvis i T4m10 översätts meningen grammatiskt korrekt, fast det är inte så idiomatisk som referensöversättningen. *Register* betecknar ord som har samma betydelse, men det valda ordet/uttrycket passar inte in i kontexten, som i T3m1: "уменьшается" är en korrekt översättning av "minskar" men när det handlar om penningvärdet då är "снижается" betydelsemässigt bättre. Med *repetition* menas det att samma eller liknande ord eller uttryck används för ofta och det är bättre att använda synonymer där – i T1m2 upprepas ordet "необходимую" exempelvis två gånger. Med *a slight inaccuracy* menas det att i princip har konstruktionen eller meningen översatts korrekt, men det finns några osäkerheter som inte påverkar begripligheten, t.ex. i T3m5 uttrycket "в состоянии" är inte exakt översättning av "har möjlighet", fast det också kan passa in i kontexten och inte påverkar den allmänna begripligheten.

Coherence visar att sammanhanget i texten är drabbat och därför finns det logiska brister. *Missing info* är en underkategori, som visar att informationen som behövs för att förstå texten har utelämnats av GT och detta påverkar begripligheten. Det fungerar ofta i samband med underkategorin *missing word*, därför att när systemet utelämnar några viktiga ord försvinner också meningen med texten. *The meaning is lost for construction/sentence* betyder att konstruktionen eller meningen är så konstiga eller ologiska att det är svårt att begripa vad som uttrycks. Den kategorin fungerar också ofta i samband med *missing info*.

Adequacy kategorin beskrivs av Daems & Macken så här:

Adequacy can be described as making sure the target text contains the same information as the source text. This means that all misinterpretations, contradictions, meaning shifts, additions or deletions are potential errors. (Daems & Macken 2013: 21).

Det finns flera underkategorier i den kategorin, men det används bara en, den viktigaste underkategorin för analysen - *terminology*. Det betecknar den varianten av MT:s översättning som inte motsvarar de generella terminologiska reglerna. Det är väldigt

viktigt att införa den underkategorin när texterna har en specifik bakgrund med många termer, för att kunna utvärdera GT:s möjligheter från olika håll.

I analysen räknas alla inkorrekta ord eller konstruktioner som ett fel även om dessa fel förekommer flera gånger. Felen som tillhör kategorierna *word order*; *disfluent sentence/construction*; *missing info*; *a slight inaccuracy*; *the meaning is lost for construction/sentence* ligger vanligtvis i längre konstruktioner och därför räknas dessa fel som ett fel per sats, inte per varje ord. Resultatet analyseras både kvantitativt, där alla fel räknas för att visa i vilka texter som mer eller mindre fel förekommer, och kvalitativt där jag ska förklara med hjälp av teori varför denna feltyp förekommer i vissa texter. Det kan tilläggas också att GT inte låter kontrollera varifrån ett konkret fel kommer, så slutsatser bara kan dras med hjälp av teoretiska kunskaper och resultat av tidigare forskning.

2.3 Resultat och analys

I Tabell 4 representeras det en generell översikt av antal ord, meningar och fel, samt förhållandet mellan fel och ord. Resultaten är jämförbara, därför att variansen i antal ord inte är så stor (206-219 ord). Samtidigt har varje text olika antal meningar, vilket beror på meningarnas olika längder. Tabellen visar, att Text 2 (medicinsk text om tuberkulos) innehåller minst fel (63) vilket utgör 28,77% av antalet ord (219) och avviker från genomsnittet (average) på 8,68%. Text 4 (tidningsartikel) har det största antalet fel (93) vilket bildar 43,06% fel av antalet ord (216). Text 3 om inflation ligger närmast genomsnittet (37,68%) medan Text 1 om byggnation har 40,29% fel av textlängden och den har bara 10 fel mindre än Text 4.

Tabell 4. Generella resultat för GT-översättningen

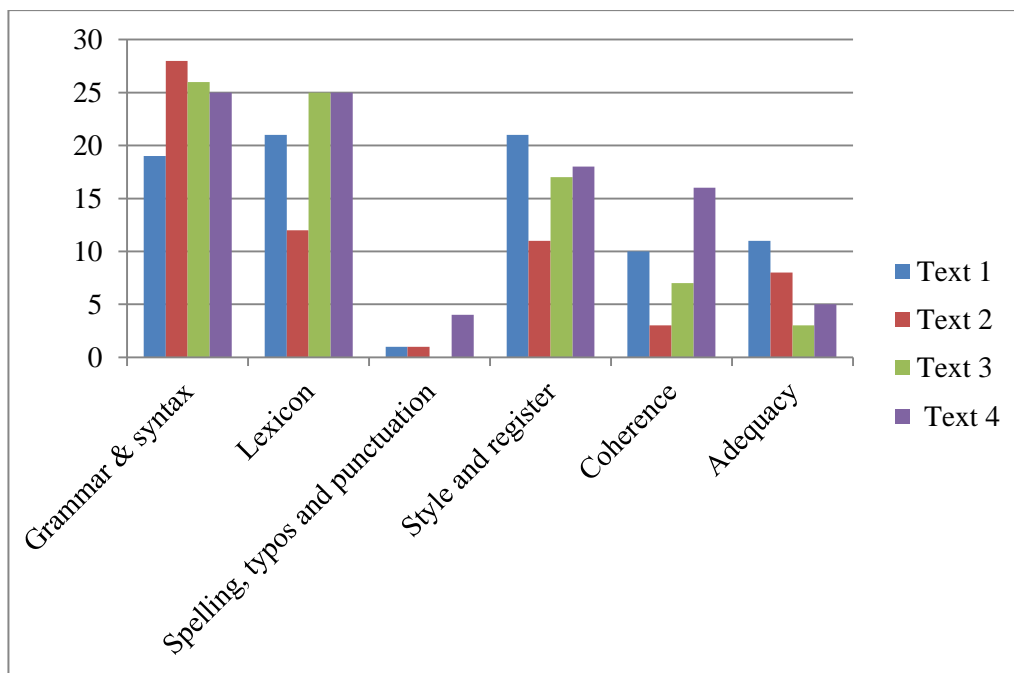
	Text 1	Text 2	Text 3	Text 4	Average
Number of words originally	206	219	207	216	212
Number of sentences	11	12	14	15	13
Number of errors	83	63	78	93	79,25
Errors / words ratio	40,29%	28,77%	37,68%	43,06%	37,45%

Tabell 5 representerar det totala resultatet av översättningen via GT. Här finns det alla felkategorierna (category) och deras underkategorier (subcategory), samt antal fel i varje underkategori vid olika sorters texter (Text 1 – Text 4).

Tabell 5. Översikt av antal fel i kategorierna

Category	Subcategory	Text 1	Text 2	Text 3	Text 4
Grammar & syntax	incorrect case	9	16	11	5
	word order	1	1	1	3
	verb form	1	2	5	5
	subject-verb agreement	4	4	0	4
	other incorrect endings	0	1	1	1
	missing constituent or preposition	4	4	8	7
	TOTAL	19	28	26	25
Lexicon	incorrect word	7	4	4	9
	missing word	9	7	10	8
	wrong preposition	2	0	1	2
	wrong word class	3	1	7	3
	extra word	0	0	3	3
	TOTAL	21	12	25	25
Spelling, typos and punctuation	capitalization	1	0	0	1
	punctuation	0	1	0	3
	TOTAL	1	1	0	4
Style and register	disfluent sentence/construction	3	4	3	3
	register	12	5	7	10
	repetition	1	1	0	1
	a slight inaccuracy	5	1	7	4
	TOTAL	21	11	17	18
Coherence	missing info	7	1	3	7
	the meaning is lost for construction/sentence	3	2	4	9
	TOTAL	10	3	7	16
Adequacy	terminology	11	8	3	5
	TOTAL	11	8	3	5
TOTAL		83	63	78	93

Sammanfattningsvis presenteras analysens resultat i Figur 2:



Figur 2. Antalet GT:s fel i de olika kategorierna i fyra texter

Grammar & syntax

I den kategorin förekommer det flest fel, vilket har varit förutsebart för mig, därför att rysk grammatik är ganska svår även för dem som har ryska som modersmål. Tabell 5 visar att det finns många fel i underkategorin *incorrect case*, nämligen i Text 2 fanns det även 16 sådana fel och i Text 3 11 kasusfel. Därigenom kan det sammanfattas att för Google Translate är det problematiskt att hantera kasusändelser, vilket orsakas av språkliga skillnader mellan språken. Som sagt tidigare, har svenskan bara två-tre kasus, medan i ryskan finns det sex olika kasus och därför GT inte kan identifiera när korrekta kasus ska användas. Å andra sidan kan det dras en annan slutsats angående skälet till felen när man tittar på några exempel av texter. Följande exempel är från T2m10:

GT: Основной причиной туберкулеза, микобактерий туберкулеза, является неподвижным, аэробная бактерия, что делит было 16-20 часов, что очень медленная скорость шаг по сравнению со многими другими бактериями, которые часто делится один раз в час или более.

De felaktiga ändelserna i GT-översättningen står i kursiv och man kan notera att det finns 7 kasusfel i en mening. Orsaken till detta kan vara meningens svåra syntaktiska

struktur, som stör semantisk förståelse och bestämning av kontexten för GT, vilket gör att systemet blandar ihop ändelser av olika kasus. I den här meningen uppfattade GT att ”primära orsaken till tuberkulos” är ”aerob bakterie”, fast huvudordet är ”Mycobacterium tuberculosis” som bestämmer kasusändelserna före och efter sig i den ryska varianten. Samma problem förekommer i T4m5 eller T1m11.5, där GT inte kan identifiera kontexten eller det finns svåra lexikala konstruktioner i texten (t.ex. sammansättningar) som är orsaken till att kasusändelserna också drabbas. Samtidigt innehåller Text 4 bara 5 kasusrelaterade fel, vilket beror på att GT har valt att utelämna stora delar av informationen från källtexten och därigenom förenklat översättningen för sig själv eller översatt satserna helt inkorrekt, så det uppträder också många andra slags fel i denna texttyp.

Å andra sidan har GT inte så stora problem med orföljden och det finns bara ett sådant fel i varje av de tre första texterna, medan Text 4 har 3 fel i ordföljden, då GT väljer att översätta svenskans ordföljd direkt när orföljden avviker från SVO-ordföljden. Detta fel förekommer oftare i just den sista texten (t.ex. mening 7). Samma resultat beskrivs i undersökningen av Aleman Tennell (2013). Därför att SMT innehåller olika MT-modeller inkl. direkt översättningsmetod och transfermetod, kan det också förekomma olika strategier för översättningen. Direkt översättning är den enklaste modellen och enligt Vauquois pyramid saknas det analysen i processen av direkta översättningar, därför kan GT välja att översätta meningen enkelt, ord för ord, vilket kanske beror på brist på nödvändig språklig information i dess korpusar.

Direkt översättningar gäller också underkategorin *verb form*. T3m5 har en relativt krånglig syntaktisk konstruktion, som innehåller en bisats samt olika objekt, vilket resulterar i direkt översättning av bl.a. verbets form, medan det behövs en annan form av verbet i den ryska kontexten (t.ex. *skall övergå* översätts som *должен превратиться*, fast det måste vara åtminstone *превратился бы*). I Text 4 finns det några meningar med vardaglig tal som delvis påverkar översättningen av verb. I mening 3 finns det utelämnning av hjälpverbet ”har” vid verbet ”försökt” och den översatta verbformen har blivit ”пытаясь” istället för ”пытались”. I Text 3 finns det inga problem med sambandet mellan subjekt och verb, i de andra texterna finns det några fel, när GT väljer pluralis för verbet, fast subjektet står i singularis eller tvärtom. Det kan

förklaras bara med det faktum att GT känner inte igen kontexten. Liknande resultat har fått Syahrina (2011) där författaren skriver att GT också förväxlar pluralis med singularis. Slutligen förekommer ganska många fel i underkategorin *missing constituent or preposition*. Dessutom har det visat sig att GT brukar utelämna några viktiga konjunktioner eller delar av dessa, precis som i undersökningen av Syahrina (2011). Den första texten har drabbats minst av olika slags fel, därför att den mestadels består av uppräknings utan några svåra grammatiska företeelser.

Lexicon

I denna kategori förekommer det också ganska många fel, speciellt i Texter 3 och 4, medan medicinsk Text 2 har avsevärt mindre fel i den kategorin. Underkategorin *incorrect ord* visar att GT översätter några ord felaktigt i varje text, speciellt i den första och den sista texterna. I Text 4 är orsaken till valet av felaktiga ord klar: det är svårt för GT att fatta vad texten mestadels handlar om, därför väljer det inkorrekt alternativ till några ord. När det gäller den första texten då finns det en tendens till att inte koppla några ord till den allmänna kontexten och alternativ ser igen felaktigt ut (t.ex. i mening 4 och 5 översätts redovisning som ”признание”, vilket är omotiverad av kontexten). I undersökningen av Hiire (2013) på språkparet svenska-estniska förekommer det också ganska många felaktiga ordöversättningar samt flera extra ord förekommer i översatta versioner, fast i denna analys inte är begripligheten drabbat av underkategorin *extra word* och det fanns bara några sådana fel.

Det finns också en annan tendens hos GT – det brukar förkorta måltexterna och detta visas genom underkategorin *missing word*, därtill utelämnas det i varje text nästan lika många ord (7-10). Det beror ofta på att några termer eller sammansättningar inte har enordiga analoger i ryskan och de ska översättas med hjälpen av flera ord (som t.ex. ”dimensionering” måste bli ”определение размеров”) medan GT översätter sådana sammansättningar eller termer ofullständigt. Allansson (2014) beskriver i sin undersökning också att översättning av sammansättningar vid språkparet engelska-svenska är ganska inkonsekvent. Dessutom, fast det är vanligt för ryskan att utelämna kopulaverbet (svenskans ”är”), är det ibland nödvändigt i satsen därför att meningen inte är idiomatiskt korrekt utan det (exemplet är från T2m3) men GT föredrar att snarare utelämna kopulaverbet ganska ofta. När det gäller prepositioner, så förekommer det

ganska många fel i språkparet engelska-svenska i uppsatsen av Askarieh (2014), fast i den felanalysen förekommer det bara små oviktiga prepositionsfel, som inte påverkar begripligheten så mycket. Underkategorin *wrong word class* har större inverkan på Text 3, där det även finns 7 fel. Det är problematiskt att bestämma varför det just är den ekonomiska texten som drabbats mest av sådana fel, därför att det inte finns några andra undersökningar med samma underkategori. Ytterligare kan det påpekas hur GT översätter avkortningar. Det finns nämligen tre av dem: den första är i T1m6 (EKS), den andra är i T2m9 (BCG-vaccin), den tredje är i T4m11 (EU). I princip har alla avkortningarna översatts korrekt, men den första har ingen motsvarighet i ryskan, det kan stanna kvar som det står på svenska efter översättningen, fast GT bara har transkriberat detta (ЭКС).

Spelling, typos and punctuation

I denna kategori finns det bara några fel, som inte påverkar begripligheten så mycket. Google Translate hanterar mycket bra rysk kommatering och har bara några problem med direkt tal i Text 4 och överför de svenska reglerna av kommatering vid direkt tal till ryskan. Exempelvis står det i T4m15 en komma efter repliken, fast enligt ryskans regler måste det stå ett tankstreck efter direkt tal. Det är intressant att vid översättning från svenska till estniska förekommer det mer fel vid kommatering enligt Hiies (2013) undersökning och det finns till och med 8 fel per 3 texter i hennes analys, medan det bara finns 4 fel per 4 texter i den här analysen. Å andra sidan utelämnar Syahrina (2011) samma underkategori vid språkparet engelska-svenska, därför att det inte finns några fel med punktuation eller tryckfel, precis som i den här analysen. När det gäller underkategorin *capitalization*, fanns det också minimala fel. Det är kanske bättre att testa den underkategorin på texter med geografisk bakgrund där det finns mer olika namn och benämningar.

Style and register

Figur 2 visar att många stilistiska fel förekommer i Text 1, vilket främst påverkas av underkategorin *register*. Nämligen finns det många ord som inte har översatts lagom precis. Det finns några ord som upprepas flera gånger i texten (t.ex. ”anvisningar”, ”redovisning”) och därför återspeglas det många fel i denna underkategori. I Text 4

finns det också flera *register* fel. Det oftaste skälet för sådana fel är att GT inte känner till textens stil och väljer ett slumpmässigt alternativ. I ryskan finns det ganska många synonymer för nästan varje ord, ibland ännu fler än i svenskan. Den språkliga orsaken förvirrar GT:s inre system och därför kan ordens betydelser vara grova. Exempelvis i T4m7 har ordet ”missa” översatts av GT som ”пропустить”, som kan vara korrekt i en annan kontext. När man pratar om ”köptillfällen” då ska det användas tillsammans med verbet ”упустить”. Det kan också tilläggas att för språkparet engelska-svenska är ordförrådet ganska stort enligt undersökningen av Syahrina (2011). Det är svårt att kontrollera om ordförrådet för språkparet svenska-ryska är lagom omfattande på exemplet av bara fyra texter men det kan konstateras att den valda motsvarigheten för översättningen av några ord inte alltid är tillräckligt precis. Det fanns mindre *register* fel i Text 2 och 3, fastän ordförrådet är svårare för uppfattning på svenska i båda texterna. Det kan bero på att de medicinska och ekonomiska texterna snarare har ett terminologiskt språk, som inte forutsätter så många varianter för översättningen.

I underkategorin *disfluent sentence/construction* förekommer det nästan samma antal fel i varje text. Det är ganska naturligt att GT inte översätter några meningar precis på samma sätt som det står i referensöversättningen. Sådana meningar är grammatiskt korrekta, fast de ska översättas mer idiomatiskt. Exempelvis översätter GT i T2m8:

GT: Собственная иммунная система человека не может нейтрализовать бугорок, но вместо этого формирует макрофагов и лимфоцитов воспалительного очага в легких, известных как гранулемы.

Man kan lägga märke till att ordöljden i den understrykna frasen är korrekt, fastän lite onaturlig för ryskan, vilket beror på den omvända ordföljden i den svenska bisatsen. GT väljer att direktöversätta detta, därför att konstruktionen verkar vara svår för systemet. Eller i T3m3: ”При увеличении денежной массы резко уменьшается правило денег.” Här kan man också se, att GT översätter den syntaktiska konstruktionen direkt, vilket överensstämmer med grammatiska regler, men det är inte en lagom idiomatisk översättning. Underkategorin *a slight inaccuracy* är lite subjektiv, det kan handla om enskilda ord eller hela konstruktioner. Exempelvis översätter GT ”immunförsvar” i Text 2 som ”иммунная система” som i princip är korrekt och passar kontexten bra, fast i

referensöversättningen står det ”иммунитет”, som kanske beskriver det svenska ordet mer exakt.

Ibland kan det vara svårt att förklara genom lingvistiska företeelser varför stilistiska fel förekommer. I praktiken har det bevisats att stilistiska fel kan påverkas av tekniska aspekter, som kallas för ”domain flexibility” (diskuterad tidigare i kapitel 1.3). Därför är det klart varför GT gör färre stilistiska fel i texter med en konkret vetenskaplig bakgrund, där avvikelserna i stilen är minimala (Text 2), fast i texttyper som är relaterade till byggnadsområde (Text 1), ekonomisk artikel (Text 3) och webbartikel (Text 4) kan det förekomma stilar som GT:s databas inte känner igen eller blandar ihop och därför förekommer det fler fel.

Coherence

Text 4 har mest fel i den kategorin. Detta orsakar först och främst underkategorin *the meaning is lost for construction/sentence* som kan också påverka textens allmänna begriplighet – ju fler fel det finns i den underkategorin desto obegripligare är texten. Text 4 kontrasterar mot alla andra texter och har 9 meningar, där det är svårt att begripa innebörden p.g.a. talrika fel, vilka bildar 60% av antalet meningar (15). Exemplet är från mening 13, där GT översätter: ”Мы получаем много жалоб от людей, которые на сети и получить разорвал.” Denna korta mening har drabbats av fel från flera kategorier vilka är *missing word*; *missing info*; *incorrect word (2)*; *verb form*; *preposition* vilket resulterar i att satsen förlorar sin mening. Det kan också bero på att maskiner saknar intuitiv uppfattning och har det svårt att adaptera semantiska nyanser.

I Text 1 finns det ett stort antal fel i underkategorin ”missing info”, som är direkt förknippat till den andra underkategorin *missing word*, därför att, som sagt tidigare, viktiga sammansättningar översatts ofullständigt, vilket i sin tur leder till brist på information. Text 2 översätts mer precis och det har inte utelämnats så mycket information från texten och detsamma gäller Text 3, där det allmänna sammanhanget av texten behålls bättre än i Text 1 och 4.

Adequacy

Den kategorin representeras genom bara en underkategori som kallas för *terminology*. Det är viktigt att jämföra hur kunnig GT är på att översätta olika termer, speciellt i de tre första texterna, som har mera specifika bakgrunder med en svår terminologi jämfört med den sista. Text 4 ska inte behandlas i den delen av analysen, trots att det finns några termer där också. Figur 2 visar att det förekommer minst fel i den ekonomiska texten, medan Text 1 har det största antalet fel. Det är lättare att följa resultatet när det visas procentuellt, genom förhållandet antalet termer och antalet felaktiga termer. Resultaten av de tre första texterna finns i tabell 6.

Tabell 6. Google Translate:s resultat i översättning av terminologi

	Text 1	Text 2	Text 3	Average
Number of words	206	219	207	211
Number of sentences	11	12	14	12
Number of incorrect terms	11	8	3	7,3
Number of terms	20	58	36	38
Right terms ratio	45,00%	86,21%	91,67%	80,70%

Tabellen visar att Text 3 har det största antalet korrekta översättningar inom fackspråket med resultatet 93,94% av korrekta termer, medan Text 1, där det finns korrekta översättningar av 45%, inte har ett lika högt resultat. Resultatet av Text 2 är också ganska högt, d.v.s. 86,21%, vilket bevisar att GT översätter medicinska termer mycket bra. Anledningen till att GT översätter olika byggnadstermer sämre än de andra kan vara att temat är ganska snävt, och även om Google Translate:s orförråd är lagom stort, finns det inte tillräckligt med material för översättning av sådana termer inom GT:s korpusar. Exempelvis har ”branddimensionering” i T4m5 översatts som ”дизайн пожарной”, och ”beständighet” som ”сопротивления”, vilket inte motsvarar referensöversättningen. Alltså hanterar systemet specifika byggnadstermerna mediokert.

Samtidigt är det anmärkningsvärt att GT översätter några svåra termer perfekt i den medicinska texten, som t.ex. ”bakterieodling” i T2m6 samt olika organrelaterade termer (”centrala nervsystemet”, ”könsorganen” i M2). Å andra sidan har undersökningen av

Patil & Davies (2014) visat att medicinska fraser översätts korrekt från engelska till de östeuropeiska språken vid bara 62% fallen. Det kan inte jämföras direkt med min undersökning därför att språkparen är olika, men ändå är det värt att kontrollera hur GT hanterar det medicinska fackspråket i andra forskningar. Ekonomiska termer översätts nästan perfekt i GT. Systemet kan hantera svåra termer som ”konsumentprisindex”, ”åtstrammingspolitik” eller ”penningmängden”. Alltså innehåller GT:s modeller och korpusar lagom med information för översättning av sådana termer.

Sammanfattningsvis är resultaten för varje text i varje kategori ganska olika. Exempelvis har Text 1 färre problem med grammatiska företeelser än de andra texterna, fast det förekommer det största antalet fel i stilen. Dessutom är GT kunnig på att översätta terminologi i ekonomisk Text 3, fast man måste vara försiktig med enklare lexikologiska element och grammatiken, där det ändå kan uppträda många fel. Text 2 har inte så många fel i stilistiska och lexikala element, men grammatiska företeelser kan vara svåra för GT att översätta, fast den typen av text (medicinsk) har färre olika fel än alla andra texterna. Den fjärde texten skiljer sig mycket från de tre andra. Det är en text av enklare journalistiskt slag med mera vardagliga språkliga konstruktioner. Den är lättare att förstå på svenska, men det finns många obegripliga meningar, lexikala och grammatiska problem och ofullständig information efter översättning till ryska, som gör texten nästan omöjlig att uppfatta. Anledningar till fel kan vara både språkliga och tekniska. Google Translate klarar sig bra med punktuation och prepositioner och har ett högt allmänt resultat vid översättning av termer i olika texter.

2.4 Sammanfattande diskussion

Undersökningen visar att Google Translate gör minst fel vid översättningen av den medicinska texten och man kan lätt förstå innebörden av den översatta texten. Resultatet kan bero på att GT kan identifiera stilen och kontexten hos denna sorts texter lätt, därför att språket i medicinska texter vanligen innehåller många specifika termer, vilket underlättar valet bland alternativ i GT:s korpusar. Beviset på detta finns också i de lexikala och stilistiska kategorierna där det förekommer ett litet antal fel i Text 2 jämfört med de andra texterna. Nyhetsartikeln är ganska problematisk för hantering och där förekommer det största antalet fel, vilket gör texten obegriplig. Anledningen till detta

kan vara att GT har svårigheter med vardagliga språkliga konstruktioner och kan inte anpassa dessa till en konkret stil. Dessutom saknas stora delar av information i måltexten. Google Translate har det fortfarande svårt med grammatiken – Menezes och Quirk skrev redan år 2005 att grammatiska företeelser representeras dåligt vid SMT och det är fortfarande ett område som ska utvecklas vidare. Texten från byggnadsområde är också ganska problematiskt för GT och det har visat sig att temat är alltför snävt för GT för att kunna översätta termer och enskilda ord lagom korrekt. Den ekonomiska texten har det bästa resultatet i översättning av termer och det kan också påpekas att sammanhanget behålls lagom bra vid översättningen till ryska, fast några lexikala och grammatiska fel förekommer ändå.

För att analysera fel har det valts ett system av sex huvudkategorier med tjugo olika underkategorier av Daems & Macken, som har kompletterats med några underkategorier från alternativa källor samt av mig själv, utgående från språkparet och karaktären av undersökningen. Inom olika felanalys har det tidigare använts olika kategoriseringar. Exempelvis i undersökningen av Hiire (2013) vid språkparet svenska-estniska används det kategoriseringen från artikeln av Vilar et al. Den huvudsakliga betoningen ligger på lexikala fel och några grammatiska former, men ett sådant system räknar inte med kontexten och beskriver inte koherens i texten. Alltså tillåter detta system inte genomföra en fullständig analys. Systemet av kategoriseringen som används i denna forskning ger möjligheten att undersöka texter ganska noggrant, det omfattar lingvistiska huvudföreteelser och är lagom flexibel för tilläggning och utelämnning av underkategorier för att anpassa systemet till språkparet och undersökningens syfte. Dessutom har jag inte kunnat hitta undersökningar, där terminologi i olika texter har kontrollerats, därför har jag inte haft möjlighet att jämföra terminologiska aspekter med andra forskningar. Den här undersökningen kan fungera som en bas för framtida forskning där det ska analyseras terminologiska fel i olika texter.

Slutligen kan det resoneras att googleöversättningar fungerar lagom bra mellan svenska och ryska för att förstå innehållet, speciellt om texten är med en vetenskaplig bakgrund. Det har tidigare påpekats av Trujillo att MT-system snarare har anpassats till texter av icke-litterära typer och resultatet av denna undersökning bevisar det. Vid journalistiska artiklar kan det förekomma många fel som påverkar begripligheten. Texterna som har

översatts via GT-verktyg i denna analys passar bra om man vill begripa vad texten handlar om generellt, men ifall måltexten behövs för publikation eller andra liknande mål ska den editeras av en mänsklig översättare.

Dessutom har Google Translate visat sig att vara ganska kunnig på att översätta termer, speciellt ekonomiska och medicinska, och man kan använda GT fullständigt för översättning av termer från dessa områden. Under analysen har det jämförts resultat med tidigare forskningar och ibland har det hittats samma slags fel som i andra undersökningar, fastän språkparen är olika. Några av underkategorierna är nya och har behandlats i den här forskningen för första gången. Därför har jag inte kunnat jämföra resultatet hos dessa underkategorier med andra undersökningar.

Avslutningsvis vill jag påpeka att Google Translate är ett modernt och snabbt verktyg för översättning, fast man måste vara medveten om att GT inte kan hantera alla texttyper och det kan förekomma språkliga fel och obegripliga meningar i några av dem. Google Translate med dess olika företeelser vid översättning ska undersökas vidare. Till exempel är det intressant hur GT översätter termer i texter från andra vetenskapliga områden. Det ska kontrolleras också om översättning av olika slags grammatiska företeelser förbättras inom GT efter den här undersökningen. Det är också möjligt att jämföra översättningar av separata meningar både i kontexten och utan den samt försöka att översätta idiom eller andra lingvistiska företeelser. Alltså är maskinöversättning ett ganska omfattande område för undersökning. Nuförtiden kan det konstateras att Google Translate är långt borta från att vara ett universalmedel för översättning och det kan inte överträffa mänskliga översättare.

LITTERATURFÖRTECKNING

1. Aleman Tengel, Maximilian: *Översättning från ryska till svenska i Google Translate.. En undersökning av satsadverbialets, objektets, subjektets och genitivattributets placering*. Magisteruppsats, Lunds universitet 2013.
2. Allansson, Jimmy: *When grundskoleklass becomes mainstream classroom. An investigation of the translation of Swedish compound nouns into English using Google Translate*. Degree Project in Linguistics, Karlstads universitet 2013.
3. Andersson, Erik: *Grammatik från grunden: en koncentrerad svensk satslära*. Uppsala: Hallgren & Fallgren 1994.
4. Andersson, Lars-Gunnar: *Språktypologi och språksläktskap 3 uppl.* Stockholm: Skriptor 1992.
5. Andersson, Lars-Gunnar: *Språktypologi och språksläktskap 4 uppl.* Stockholm: Liber AB 2001.
6. Askarieh, Sona: *Cohesion and Comprehensibility in Swedish-English Machine Translated Texts*. Magisteruppsats, Linköping universitet 2014.
7. Barreiro, A. et al: „Linguistic evaluation of support verb constructions by OpenLogos and Google Translate“. I: *LREC 2014: Ninth International Conference on Language Resources and Evaluation*; Reykjavik, Iceland 2014 s.35-40
8. Bird, Steven et al: *Natural Language Processing with Python*. USA: O'Reilly Media 2009
9. Costa-jussà, Marta R. & Fonollosa, José A.R.: “Latest trends in hybrid machine translation and its applications“. I: *Computer Speech and Language* 32 (2015), s. 3–10
10. Daems, Joke & Macken, Lieve: *Annotation Guidelines for English-Dutch Translation Quality Assessment. Version 1.0. LT3 Technical Report - LT3 13.02*. University College Ghent, 2013

11. Daems, Joke, Macken, Lieve, & Vandepitte, Sonia: „On the origin of errors: a fine-grained analysis of MT and PE errors and their relationship“. I: Reykjavik, Iceland, Nicoletta Calzolari et al (red.): *LREC 2014: Ninth International Conference on Language Resources and Evaluation*, May 26-31, 2014, s.62-68.
12. Daems, Joke, Macken, Lieve, & Vandepitte, Sonia: „Quality as the sum of its parts: a two-step approach for the identification of translation problems and translation quality assessment for HT and MT+PE“. I: Nice, France: *Proceedings of MT Summit XIV Workshop on Post-editing Technology and Practice (WPTP-2)*, 2 September 2013, s. 63-71.
13. Ethnologue. Languages of the World. Statistical Summaries. <https://www.ethnologue.com/statistics/size> [läst 24.04.2015]
14. Ethnologue. Languages of the World. Swedish. <https://www.ethnologue.com/language/swe> [läst 24.04.2015].
15. Francis M.Tyers & Jacob Nordfalk: „Shallow-transfer rule-based machine translation for Swedish to Danish“. I: *Proceedings of the First International Workshop on Free/Open-Source Rule-Based Machine Translation*, Juan Antonio Pérez-Ortiz, Felipe Sánchez-Martínez, Francis M.Tyers (red.). Universitat d’Alacant, Alacant, Spain, 2-3 November 2009, s. 27-33
16. Google Company. Our history in depth. <http://www.google.ee/about/company/history/> [läst 22.04.2015]
17. Google Switches To Its Own Translation System. I: *Google Operating System*, den 22 oktober 2007. <http://googlesystem.blogspot.com/2007/10/google-translate-switches-to-googles.html> [läst 22.04.2015]
18. Google Translate. About. http://translate.google.com/about/intl/en_ALL/index.html [läst 24.04.2015]
19. Google Translate:s webbsida. <https://translate.google.com/> [läst i mai]
20. Goutte, Cyril et al: *Learning Machine Translation*. London: *MIT Press* 2009
21. Grammatisk Informationsportal GRAMOTA.RU. Справочно-информационный портал ГРАМОТА.РУ – русский язык для всех. <http://www.gramota.ru/> [läst i mai]
22. Groves, Michael & Mundt, Klaus: “Friend or foe? Google Translate in language for academic purposes”. I: *English for Specific Purposes* 37 (2015), s. 112-121

23. Hiire, Kertu: *Översättning från svenska till estniska med hjälp av Google Translate*. Bacheloruppsats, Tartu universitet 2013.
24. Hu, Julian: "Statistical machine translation based on translation rules". I: *Journal of Chemical and Pharmaceutical Research* 6(7) (2014), s. 1628-1635
25. Hutchins, W. John: "Machine translation: general overview". I: Mitkov, Ruslan (red.): *The Oxford Handbook of Computational Linguistics*. Oxford: University Press 2003, s. 501-511
26. Hutchins, W. John: „Machine translation: a concise history”. I: *Journal of Translation Studies. Special issue: The teaching of computer-aided translation*. 13:1-2 (2010), s. 29-70
27. Hutchins, W. John: *Early years in machine translation: memoirs and biographies of pioneers*. Amsterdam; Philadelphia: John Benjamins 2000
28. Inflation (språk: svenska). *Wikipedia. Wikimedia foundation*. <http://sv.wikipedia.org/wiki/Inflation> [läst 27.04.2015]
29. Koehn, Philipp: *Statistical machine translation*. Cambridge: Cambridge University Press 2010
30. Menezes, Arul & Quirk, Chris: "Dependency treelet translation: the convergence of statistical and example-based machine-translation?". I: *Proceedings of Second Workshop on Example-Based Machine Translation; MT Summit X*, Phuket, Thailand 2005 s.99-108
31. Olsson, Alena: *Tvåspråkiga barn berättar på ryska och svenska. En explorativ studie om narrativ förmåga hos 6- till 7 åringar*. Examensarbete, Uppsala Universitet 2013.
32. Patil, Sumant & Davies, Patrick: "Use of Google Translate in medical communication: evaluation of accuracy". I: *BMJ*, 2014. <http://www.bmj.com/content/349/bmj.g7392> [läst 27.04.2015]
33. *Så slipper du bli lurad av bluffsajter på nätet*, I: SvD Näringsliv, den 30 april 2015 http://www.svd.se/naringsliv/pengar/sa-slipper-du-bli-lurad-av-bluffsajter-pa-natet_4523838.svd [läst 30.04.2015]
34. Salimi, Jonni: *Machine translation of fictional and non-fictional texts. An examination of Google Translate's accuracy on translation of fictional versus non-fictional texts*. Bacheloruppsats, Stockholms universitet 2014.

35. Syahrina, Alvi: *Online machine translator system and result comparison – statistical machine translation vs hybrid machine translation*. Bacheloruppsats, Borås universitet 2011.
36. Trujillo, Arturo: *Translation Engines: techniques for machine translation*. London: *Springer* 1999.
37. Tuberkulos (språk: svenska). *Wikipedia. Wikimedia foundation*. <http://sv.wikipedia.org/wiki/Tuberkulos> [läst 27.04.2015]
38. Vilar, David; Xu, Jia; Fernando D'Haro, Luis; Ney, Hermann: „Error analysis of statistical machine translation output“. I: Genoa, Italy: *LREC-2006: Fifth International Conference on Language Resources and Evaluation*, 22-28 May 2006, s. 697-702

BILAGOR

Bilaga 1 Text 1.

	Text på svenska	Mänsklig översättning	Google Translate:s översättning
1	En viktig del för ett godkännande är att det ska finnas en tillhörande handling.	Важным фактором для того, чтобы получить одобрение, является наличие соответствующего документа.	Важной частью утверждения является то, что должно быть связано действие.
2	Denna ska innehålla all nödvändig information som behövs för att produkten ska kunna användas.	Он должен содержать всю информацию, необходимую для того, чтобы продукт мог быть использован.	Он должен содержать всю необходимую информацию, необходимую для продукта, который будет использоваться.
3	Se exempel på nödvändigt innehåll nedan:	Посмотрите на пример необходимой информации, приведенный ниже:	Посмотрите примеры необходимого содержания ниже:
3.3	- Produktbeskrivning (tvärsnitt, diameter, lindningsavstånd, längd mm.)	- Описание продукта (поперечное сечение, диаметр, шаг обмотки, длина и проч.)	- Описание продукта (сечение, диаметр, расстояние обмотки, длина и т.д.).
3.4	- Förutsättningar för användning	- Предпосылки для использования	- Условия использования
3.5	- Deklarerade värden på relevanta egenskaper	- Сообщенные значения соответствующих свойств	- Объявленные значения соответствующих свойств
3.6	- Anvisningar för hantering, transport, lagring mm.	- Руководство по обращению, транспортировке, хранению и проч.	- Инструкция по обработке, транспортировке, хранению и т.д
3.7	- Anvisningar för installation	- Руководство по установке	- Инструкция по установке
3.8	- Projekteringsanvisningar	- Руководство по проектированию	- Указания по проектированию
4	Redovisning bör finnas för hur statiska konstruktionsberäkningar utförs vid användande av glasfiberarmering i betongkonstruktioner.	Отчет должен касаться того, как производятся статические расчеты для конструкций при использовании армирования из стекловолокна в бетонных конструкциях.	Признание должно быть доступно для статических проектных расчетов, выполненных с использованием арматуры из стекловолокна в бетонных конструкциях.

5	Redovisning bör finnas för hur branddimensionering utförs vid användande av glasfiberarmering i betongkonstruktioner.	Отчет должен касаться того, как производится определение размеров противопожарной системы при использовании армирования из стекловолокна в бетонных конструкциях.	Признание должны быть доступны для дизайн пожарной выполняется с использованием арматуры из стекловолокна в бетонных конструкциях.
6	Dimensionering ska utföras enligt tillämpliga eurokoder tillsammans med nationella val i EKS.	Определение размеров должно производиться в соответствии с применяемыми еврокодами вместе с национальным стандартом в системе EKS.	Размеров должны быть выполнены в соответствии с действующими Еврокодах совместно с национальными выборами в ЭКС.
7	Notera att samtliga relevanta egenskaper ska vara verifierade genom t.ex. provning.	Обратите внимание на то, что все соответствующие свойства должны быть подтверждены, например, в ходе испытаний.	Обратите внимание, что все соответствующие характеристики должны быть проверены на примере Тестирование.
8	Provningar ska vara utförda av ackrediterade testlab.	Испытания должны быть осуществлены аккредитованной тестовой лабораторией.	Испытания должны проводиться в аккредитованной лаборатории.
9	Hänvisningar till litteratur kan inte användas som underlag för verifiering av egenskaper.	Ссылки на литературу не могут быть использованы в качестве основы для подтверждения свойств.	Ссылки на литературные источники не могут быть использованы в качестве основы для проверки свойств.
10	Jag bifogar återigen de generella kraven som gäller för ett godkännande:	Я вновь прилагаю общие требования, которые необходимы для получения одобрения:	Я снова приложить общие требования, предъявляемые к утверждению:
10.1	- Beskrivning av produkten och dess avsedda användning	- Описание продукта и его соответствующего применения	- Описание продукта и его предполагаемого использования
11	Bedömningsunderlag:	Основания для вынесения оценки:	Оценка:
11.1	- Materialspecifikationer	- Спецификации материалов	- Спецификации материалов
11.2	- Deklarering av innehåll av farliga ämnen	- Сообщение о содержании опасных веществ	- Декларация о наличии опасных веществ

11.3	- Verifiering av alla relevanta egenskaper (t.ex. provningsrapporter)	- Подтверждение всех соответствующих свойств (например, рапорт о результатах испытаний)	- Проверка всех соответствующих характеристик (например, отчеты испытаний)
11.4	- Verifiering av beständighet	- Подтверждение прочности	- Проверка сопротивления
11.5	- Tillhörande handlingar (t.ex. dimensionerings- och projekteringsanvisningar, installationsanvisningar, drifts- och skötselansvisningar).	- Прилагаемые документы (например, руководство по определению размеров и проектированию, руководство по установке, руководство по эксплуатации и уходу).	- Связанные документы (например, размеров и строительных процедур, инструкций по монтажу, эксплуатации и инструкции по техническому обслуживанию).
11.6	- Dokumenterad egenkontroll för tillverkningen	- Документированный собственный контроль производства	- Документально самоконтроль производства
11.7	- Inledande kontrollbesök	- Предварительный контрольный осмотр	- Первоначальный осмотр
11.8	- Avtal för övervakande kontroll av tillverkningen inklusive kontrollinstruktioner	- Договор о надзоре за производством вместе с инструкциями по контролю	- Договоры для контроля производства, включая инструкции управления
11.9	- Föreslag för märkning av produkten	- Предложение о маркировке продукта	- Рекомендуются для маркировки продукции

Bilaga 1 Text 2.

	Text på svenska	Mänsklig översättning	Google Translate:s översättning
1	Tuberkulos är en infektionssjukdom som vanligtvis orsakas av tuberkelbakterien <i>Mycobacterium tuberculosis</i> .	Туберкулёз является инфекционным заболеванием, которое, как правило, вызвано бактерией группы <i>Mycobacterium tuberculosis</i> .	Туберкулез является инфекционным заболеванием, как правило, вызваны бугорок микобактерий туберкулеза.
2	Tuberkulos drabbar framförallt lungorna, men kan även angripa andra organ som centrala nervsystemet, könsorganen, blodomloppet, skelettet, leder eller huden.	Туберкулёз поражает, прежде всего, легкие, но также может атаковать и другие органы, такие как: центральную нервную систему, половые органы, систему кровообращения, скелет, суставы или кожу.	Туберкулез в основном поражает легкие, но также может атаковать другие органы, центральную нервную систему, половые органы, кровь, кости, суставы или кожи.
3	De klassiska symtomen på tuberkulos är kronisk hosta med blodigt slem, feber, nattlig svettning och vikt förlust.	Классическими симптомами туберкулеза являются хронический кашель с кровавой мокротой, лихорадка, ночная потливость и потеря веса.	Классические симптомы туберкулеза хронический кашель с кровавой мокротой, лихорадка, ночные поты и потеря веса.
4	Infektion i andra organ kan ge en rad olika symtom beroende på vilket organ som är drabbat.	Заражение других органов может вызвать ряд различных симптомов в зависимости от того, какой орган поражён.	Заражение других органов может вызвать различные симптомы в зависимости от которых зависит органом.
5	Medicinsk diagnos av tuberkulos baseras främst på röntgenbilder av lungorna och tuberkulintest på huden.	Медицинский диагноз туберкулеза основывается главным образом на рентгеновских снимках легких и на кожной туберкулиновой пробе.	Медицинская диагностика туберкулеза основана главным образом на рентгеновских снимках легких и туберкулиновой кожи.
6	Mikroskopundersökning och bakterieodling av kroppsvätskor kan också användas.	Микроскопическое исследование и бактериальная культура, полученная из жидкостей организма, также могут быть использованы.	Микроскопическое исследование и бактериальная культура жидкостей организма также могут быть использованы.

7	Tuberkulosbehandling en är långvarig och omständlig och kräver många månaders intag av olika sorters antibiotika.	Лечение туберкулеза является длительным и подробным и требует многих месяцев приема различных антибиотиков.	Лечение туберкулеза является длительным и утомительным и требует многих месяцев приема различных антибиотиков.
8	Människans eget immunförsvar kan inte oskadliggöra tuberkelbakterien utan istället bildar makrofager och lymfocyter inflammatoriska härdar i lungorna som går under namnet granulom.	Собственный иммунитет человека не может нейтрализовать бактерию, вызывающую туберкулёз, но вместо этого макрофаги и лимфоциты образуют воспалительные очаги, под названием гранулёмы, в лёгких.	Собственная иммунная система человека не может нейтрализовать бугорок, но вместо этого формирует макрофагов и лимфоцитов воспалительного очаги в легких, известных как гранулемы.
9	Tuberkulos förebyggs genom screeningtester och vaccinering, oftast med BCG-vaccin.	Туберкулез предотвращается с помощью скрининговых тестов и вакцинации, чаще с помощью вакцины БЦЖ.	Туберкулез можно предотвратить с помощью скрининговых тестов и вакцинации, как правило, с вакциной БЦЖ.
10	Den primära orsaken till tuberkulos, Mycobacterium tuberculosis, är en orörlig, aerob bakterie som delar sig var 16–20 timme, vilket är en mycket långsam delningstakt jämfört med många andra bakterier som ofta delar sig en gång i timmen eller mer.	Основная причина туберкулеза, Mycobacterium tuberculosis, является неподвижной аэробной бактерией, которая делится каждые 16-20 часов, что является очень медленным темпом размножения по сравнению со многими другими бактериями, которые часто делятся один раз в час или чаще.	Основной причиной туберкулеза, микобактерий туберкулеза, является неподвижным, аэробная бактерия, что делит было 16-20 часов, что очень медленная скорость шаг по сравнению со многими другими бактериями, которые часто делятся один раз в час или более.
11	M. tuberculosis tillhör släktet mykobakterier och de har ett yttre membran som är uppbyggt av ett dubbelskikt av lipider.	M. tuberculosis относится к роду микобактерий и имеет наружную мембрану, состоящую из двойного слоя липидов.	M. туберкулез принадлежит к роду микобактерий и имеют наружную мембрану, состоящую из двойного слоя липидов.
12	Därför fungerar gramfärgning dåligt på M. tuberculosis, då den antingen bara färgas svagt eller inte håller kvar färgen (vilket beror på de lipider och den mykolsyra cellen är uppbyggd av).	Поэтому окрашивание по Граму плохо работает в отношении M. tuberculosis, которая или просто слабо окрашивается, или не сохраняет цвет (что зависит от липидов и миколовой кислоты, из которых построена клетка).	Поэтому работает Граму плохо на микобактерии туберкулеза, то она либо просто окрашивают слабо или не сохраняют цвет (за счет липидов и клеточных миколовых кислоты состоит из).

Bilaga 3 Text 3.

	Text på svenska	Mänsklig översättning	Google Translate:s översättning
1	Med inflation menas vanligtvis att penningvärdet minskar.	Под инфляцией обычно подразумевается снижение стоимости денег.	Учитывая, что инфляция, как правило, означает, что стоимость денег уменьшается.
2	Den ursprungliga definitionen är "ökning av penningmängden".	Изначальное определение звучит как "увеличение денежной массы."	Оригинальный определение "увеличение денежной массы."
3	När penningmängden ökar kraftigt så minskar i regel penningvärdet.	При резком увеличении денежной массы, как правило, стоимость денег падает.	При увеличении денежной массы резко уменьшается правило денег.
4	En effekt av inflation är att den allmänna prisnivån stiger, varför inflation ofta mäts med konsumentprisindex och uppfattas av många som synonymt till detta.	Одним из эффектов инфляции является повышение общего уровня цен, поэтому инфляция часто измеряется с помощью индекса потребительских цен и воспринимается многими как синоним того.	Влияние инфляции является то, что общий уровень цен растет, почему инфляция часто измеряется с помощью индекса потребительских цен и воспринимается многими как синоним для этого.
5	För att en inflationsimpuls skall övergå i stadigvarande inflation måste den eller de som har möjlighet att skapa mer pengar, vanligtvis centralbanken och vanliga banker, öka penningmängden.	Для того, чтобы инфляционный импульс перешёл в постоянную инфляцию, те, кто имеют возможность создавать больше денег (обычно это центральный банк и обычные банки), должны увеличить денежную массу.	Для инфляционный импульс должен превратиться в постоянного инфляции есть, или те, кто в состоянии создать больше денег, как правило, центральный банк и регулярные банки, увеличивая денежную массу.
6	Penningmängden är ett mått på de likvida medel som cirkulerar i ekonomin.	Денежная масса является мерой ликвидных средств, циркулирующих в экономике.	Денежная масса является мерой наличных, циркулирующих в экономике.
7	Inflationens verkningar (förutom att både löner och priser stiger) kan innebära en omfördelning av både förmögenheter och inkomster, då inflationen är oförutsedd.	Последствия инфляции (помимо роста как зарплат, так и цен) могут заключать в себе перераспределение как богатства, так и доходов, если инфляция непредвиденная.	Последствия инфляции (кроме того, что оба заработной платы и рост цен) может предусматривать перераспределение богатство и доход, когда инфляция неожиданно.

8	Dessutom kan utrikeshandeln påverkas av höjda prisnivåer, särskilt vid fast växelkurs.	Кроме того, внешняя торговля может зависеть от высокого уровня цен, особенно в случае фиксированного обменного курса.	Кроме того, внешняя торговля влияет на более высоких уровнях цен, особенно на фиксированный обменный курс.
9	Inflationsbekämpning drivs genom åtstramningspolitik för att dämpa efterfrågan.	Борьба с инфляцией проводится с помощью политики жёсткой экономии для того, чтобы понизить спрос.	Борьба с инфляцией, приводимый в политике жёсткой экономии, чтобы обуздать спрос.
10	Det är främst genom en åtstramad finanspolitik (höjda skatter och minskade offentliga utgifter) och penningpolitik (höjda räntor), som staten kan dämpa inflationen.	Это осуществляется, в первую очередь, через ужесточённую налогово-бюджетную политику (повышенные налоги и урезанные государственные расходы) и денежно-кредитную политику (более высокие процентные ставки), с помощью которых государство может обуздать инфляцию.	Это, главным образом, ужесточение налогово-бюджетной политики (повышение налогов и государственных расходов) и денежно-кредитной политики (более высокие интрессы), которые государство может обуздать инфляцию.
11	Det är viktigt att en inflationsbekämpande politik görs trovärdig så att inflationsförväntningarna minskar.	Важно, чтобы антиинфляционная политика проводилась добросовестно, чтобы инфляционные ожидания снижались.	Важно, что антиинфляционная политика заслуживает доверия, так что инфляционные ожидания снижаются.
12	Om inflationen är hög kan människor börja samla på sig saker, eftersom alternativet - att inte köpa - minskar värdet av det man har.	Если инфляция высокая, люди могут начать накапливать вещи, потому что альтернатива - не покупать - снижает ценность того, что есть у человека.	Если инфляция высокая, люди начинают накапливать вещи, потому что альтернатива - не покупать - уменьшает значение, что у вас есть.
13	Detta är en ur samhällssynpunkt oattraktiv utveckling, eftersom investeringarna sker i varor istället för produktion.	С общественной точки зрения – это непривлекательное развитие, поскольку инвестиции вкладываются в товары, а не в производство.	Это социально непривлекательно развития, поскольку инвестиции осуществляются в товарах, а не производства.
14	Detta kan även leda till bubblor.	Это даже может привести к образованию экономических пузырей.	Это также может привести к образованию пузырьков.

Bilaga 4 Text 4.

	Text på svenska	Mänsklig översättning	Google Translate:s översättning
1	86 procent av svenskarna handlar på internet, en fördubbling mot för tio år sedan.	86 процентов шведов делают покупки в Интернете, что в два раза больше, чем десять лет назад.	86 процентов шведов покупки онлайн, удвоить ставку десять лет назад.
2	Det visar sig att många är missnöjda med sina köp, men trots detta är det få som klagar.	Оказывается, многие недовольны своими покупкам, но, несмотря на это, тех, кто жалуется - мало.	Оказывается, что многие недовольны своей покупкой, но, несмотря на это, мало жалуются.
3	– Hälften av dessa, omkring 180 000 av de missnöjda konsumenterna, inte försökt åtgärda felet, säger Gunnar Larsson.	– Половина из них, около 180 000 из недовольных потребителей, не пытались исправить ошибку – говорит Гуннар Ларссон.	- Половина из них, около 180 000 из недовольных потребителей, не пытаюсь исправить ошибку, говорит Гуннар Ларссон.
4	Hans uppmaning till alla konsumenterna är att bli bättre på att reklamera.	Его призыв ко все потребителям - будьте готовы предъявлять претензии.	Его призыв ко всем потребителям, чтобы получить лучше жаловаться.
5	Den vanligaste orsaken till passiviteten är den som köpt något tycker att det handlar om för lite pengar.	Наиболее распространенной причиной пассивности является то, что, покупая что-либо, человек считает, что речь идет о слишком малых деньгах.	Наиболее распространенной причиной пассивности, кто купил что-то думаю, что это о слишком мало денег.
6	Det är inte mödan värt att klaga.	Проблема не стоит того, чтобы жаловаться.	Не стоит проблема, чтобы жаловаться.
7	Samtidigt riskerar svenska konsumenter att missa bra köptillfällen på grund av misstro till utländska sajter.	В то же время шведские потребители рискуют упустить хорошие возможности покупки из-за недоверия к иностранным сайтам.	В то же опасности шведских потребителей пропустить хорошие возможности для покупки из-за недоверия к иностранным сайтам.
8	– Vi är ganska försiktiga när vi handlar på sajter utanför Sverige.	– Мы довольно осторожны, когда совершаем покупки на сайтах за пределами Швеции.	- Мы вполне осторожны, когда мы по магазинам в местах за пределами Швеции.
9	Kanske för försiktiga.	Возможно, слишком осторожны.	Возможно, слишком осторожны.
10	Här finns det bra produkter till bra priser om vi vågar använda dem, säger Gunnar Larsson.	Здесь можно найти хорошие продукты по хорошим ценам, если мы осмелимся их использовать - говорит Гуннар Ларссон.	Здесь есть хорошие продукты по хорошим ценам, если мы смеем их использовать, говорит Гуннар Ларссон.

11	Konsumentverket vill nu se flera åtgärder för att förbättra förutsättningarna för köpare, bland annat skapa samma regler för konsumentskydd inom EU och förstärka Konsumentverkets möjligheter att agera mot näthandlare som inte följer regelverket	Теперь Управление защиты прав потребителя хочет рассмотреть ряд мер по улучшению условий для покупателей, в частности, создать такие же правила для защиты прав потребителя внутри ЕС и укрепить возможности Управления защиты прав потребителя действовать против интернет-магазинов, которые не соблюдают правила.	Потребительские Агентство теперь хочет, чтобы увидеть дополнительные меры по улучшению условий для покупателей, среди прочего, создать такие же правила для защиты потребителей в ЕС и укреплять способность потребителя Агентства действовать против интернет-магазинов, которые не соблюдают правила.
12	Mot rena bedrägerisajter och uppenbart oseriösa e-handlare vill Konsumentverket få nya vapen.	Против чисто мошеннических сайтов и очевидных несерьёзных интернет-магазинов Управление защиты прав потребителя хочет разработать новые методы борьбы.	На чистый обман сайтов и понятным изгоем электронной розничной торговли, как потребителей агентства получите новое оружие.
13	– Vi får många klagomål från personer som är ute på nätet och blir lurade.	– Мы получаем много жалоб от людей, которые выходят в сеть и становятся обманутыми.	- Мы получаем много жалоб от людей, которые на сети и получить разорвал.
14	Att kunna släcka ned eller åtminstone kunna pausa hemsidor så länge utredning pågår tycker vi är en bra idé.	Суметь запретить или хотя бы приостановить деятельность сайтов, пока продолжается расследование, является, по нашему мнению, хорошей идеей.	Будучи в состоянии выключить, или, по крайней мере, быть в состоянии сделать паузу сайты тех пор, как расследование продолжается, мы думаю, это хорошая идея.
15	Det bör utredas vidare, säger Gunnar Larsson.	Это должно расследоваться дальше - говорит Гуннар Ларссон.	Это должны быть исследованы дальше, говорит Гуннар Ларссон.

Bilaga 5 Översikten av andra undersökningar om översättning i Google Translate.

Språk-par	Författaren och år	Metod	Resultat
Ryska-svenska	Aleman Tennell, 2013	Analys av 14 korta textstycken och 19 exempelmeningar, för att se hur GT hanterar rysk ordföljd	Satsadverbialet <i>inte</i> placeras på fel plats vid 50% av fallen i huvudsatser efter översättningen i GT; Genitivkonstruktioner översätts fel genom att inte frångå ryskans ordföljd, men de översätts korrekt till alternativa korrekta svenska konstruktioner; SVO meningar översätts felfritt, fast vid avvikande ordföljden förekommer det fel.
Svenska-estniska	Hiire, 2013	Komparativ analys, jämförande fel analys av tre olika texter översatta via GT	Lärobokstexter och nyhetstexter är begripliga vid översättningen via GT, och har 64% respektive 63% av begriplighet, medan skönlitterära texter verkar vara svåra att förstå (27% begriplig).
Svenska-Danska	Francis & Nordfalk, 2009	Komparativ analys mellan tre MT system, inkl GT.	Fast variationen mellan resultaten var oväsentlig, hade GT mest fel i „word error rate (WER)“, fast mindre fel i „positionindependent error rate (PWER)“, jämfört med två andra system. GT översättningar skiljer sig mycket åt andra systems översättningar och har artefakter av engelska.
Engelska-svenska-engelska	Syahrina, 2011	Jämförande analys av två MT system: GT och Systran på olika textavsnitter.	GT visar bättre resultat jämfört med Systran. Emellertid gör GT många fel i pluralis och förväxlar bestämdhet hos substantiv, utelämnar konjunktioner vid översättning från engelska till svenska. Google klarar sig bra med syntaktiska strukturer och termer och har större ordförråd än Systran.

	Salimi, 2014	BLEU analys av litterära (10 stycken) och icke-litterära texter (10 stycken) översätta via GT.	BLEUs 40 poäng 40 syftar på en mycket bra översättning. Icke-litterära texter fick 32.16 poäng (varians: 20.12 – 58.08) medan litterära fick bara 27.75 (varians 17.99 – 38.45). Det bästa resultatet fick två juridiska texter (58.08). En av litterära texterna fick den lägsta poängen på 17.99.
	Allansson, 2014	Analys av översättningen av svenska sammansättningar via GT i nio texter.	Undersökningen visade att GT är ganska inkonsekvent i översättning av sammansättningar. Därför var det svårt för författaren att systematisera resultat.
	Askarieh, 2014	Fel analys av 2 texter översatta via GT och Bing.	GT hade i genomsnitt få fel, som inte påverkade begriplighet av översättningar. Fel som förekom gällde prepositioner, artiklar och genitivformer, samt förvanskades några substantiv, verb och adjektiv av den makinella översättningen.
Engelska- ... 13 europeiska, 11 asiatiska och 2 afrikanska språk	Patil & Davies, 2014	Tio medicinska fraser utvärderades i 26 olika språk, vilket resulterade i 260 översatta fraser, resultaten jämfördes med mänsklig översättning.	GT har bara 57,7% grammatisk noggrannhet när det används för översättning av medicinska fraser och ska inte användas för viktiga medicinska kommunikationer. Översättningarna till de västra-europeiska och öst-europeiska språken var mest begripliga (74% respektive 62%), de afrikanska språken hade det lägsta resultatet (45%).

RESÜMEE

Käesoleva bakalaureusetöö teemaks on “Google Translate’i tõlkimise tehnika rootsi keelest vene keelde: nelja erineva taustaga tekstide analüüs“. Vabaaja veetmiseks, suhtlemiseks ja õppimiseks kasutatakse tänapäeval üha rohkem interneti. Inimeste maksimaalne integratsioon virtuaalkeskonnas, mitmesuguste õppevõimalusi kasutamine ja palju muud on raskendatud keelebarjääri tõttu. Tõlkimine on tõhus vahend, mis aitab vähendada keelebarjääri mõju. Google Translate selle vahendi populaarse esindajana pakub kasutajatele kaasaegset, kiiret ja tasuta masintõlget. Uuringu eesmärgiks on kontrollida, kuidas tõlge toimib tõlge rootsi keelest vene keelde, millised vead tekivad tõlkimise protsessis ja kuidas on võimalik neid süstematiseerida. Samuti kontrollitakse antud bakalaureusetöös Google Translate’i oskust tõlkida erinevaid termineid. Analüüsimiseks kasutatakse nelja erinevat teksti: formaalne kiri ehituse teemal, meditsiiniline erialane tekst, majanduslik tekst ning lihtsamat tüüpi ajakirjanduslik artikkel. Kõik vead mis tekivad Google Translate’i tõlgetes grupeeritakse erinevatesse kategooriatesse, analüüsitakse ja tehakse järeldusi.

Teoreetilises osas käsitletakse masintõlke definitsiooni, põhilisi mudeleid ja arengukäiku, samuti kirjeldatakse statistilise masintõlke ja Google Translate’i olemust, tuuakse välja erinevused rootsi ja vene keele grammatika vahel ning vaadeldakse peamisi tulemusi varasemates uuringutes. Empiirilises osas presenteeritakse materjali, meetodi valikut ja analüüsi tulemusi. Arutletakse vigade esinemise üle, tehakse järeldusi.

Analüüsi käigus selgus, et Google Translate teeb kõige rohkem vigu ajakirjanduslikus tekstis, kõige vähem vigu esineb meditsiinilises tekstis. Majandusliku teksti puhul on vigade arv veidi suurem kui meditsiinilises tekstis, kuid ehitusteemalises tekstis esineb peaaegu sama palju vigu kui ajakirjanduslikus artiklis. Kõige rohkem vigu teeb Google Translate grammatikas ja leksikas: ebatäpsusi leidub eelkõige käändelõppudes ja

tegusõna vormides. Peale selle esineb tihti nii valesti tõlgitud, kui ka tõlkimata ja tekstist välja jäetud sõnu. Kohati esineb ka stilistilisi vigu.

Tõlkevääratuste peamised põhjused on tehnilise ja lingvistilise iseloomuga. Google Translate'il on raskusi konteksti äratundmisega või on selle andmebaasides korrektse tõlkimise jaoks liiga vähe informatsiooni. Olulist rolli mängivad ka keeltevahelised erinevused ning keerulised süntaktilised konstruktsioonid. Samal ajal ilmneb, et Google Translate saab oskussõnavaraga päris hästi hakkama. Antud kategooria vigade protsent oli üks madalmaid kõikide veakategooriate lõikes ning eriti majandus- ja meditsiinitekstides.

Kokkuvõtvalt võib öelda, et 2015. aasta Google Translate'i võime tõlkida rootsi keelest vene keelde on hea: tõlked on loetavad ja arusaadavad. Mõnevõrra on erandiks ajakirjanduslik tekst, mille originaal on algsetest tekstidest kõige lihtsam, kuid mille tõlge on kõige kehvem. Vajaduse korral soovitab antud töö autor kasutada Google Translate'i üksnes mitteajakirjanduslike, erialaste ja teaduslike tekstide tõlkimiseks, mida iseloomustavad täpsed süntaktilised konstruktsioonid ning selgepiiriline sõnavara ja terminoloogia.

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Aleksandra Fomicheva, (sünnikuupäev: 09.06.1993)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose
„Google Translate:s översättningsteknik från svenska till ryska: en analys av fyra översatta texter med olika bakgrunder”, mille juhendaja on Maiu Elken,
 - 1.1.reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
 - 1.2.üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 22. Mai 2015