



LUNDS
UNIVERSITET

Språk- och litteraturcentrum

Översättarutbildningen

EXAMENSARBETE VT 2012

MASTER I ÖVERSÄTTNING

Specialisering i spanska

Googleöversättning

Om Google Translate:s möjligheter och
begränsningar vid översättning till svenska

Författare:

Björn Moberg

Borm86@gmail.com

Handledare:

Lars-Johan Ekerot

Sammandrag

I den här uppsatsen undersöks maskinöversättningstjänsten Google Translates förtjänster och begränsningar vid översättning till svenska. Detta görs genom att närstudera ett litet material av googleöversatt text till svenska från engelska och spanska och sedan med denna närstudie som grund laborera med onlinetjänsten för att belysa hur översättningarna kan variera beroende på form och kontext. Google Translates problemområden delas i denna uppsats in i problem som berör korrekthet respektive begriplighet. Med problem som berör korrekthet åsyftas problem som gör översättningarna ogrammatiska utan att det går ut över textinnehållet. Dessa problem övergår i problem med begriplighet om de ändrar eller otydliggör texterna innehållsmässigt. En viktig slutsats av studien är att Google Translate tycks klara av konventionaliserade standardfraser bättre än mer oväntade (statistiskt mindre förekommande) fraser.

Nyckelord: Google Translate, statistisk maskinöversättning

Engelsk titel: *Google translations: A study of Google Translate's possibilities and limitations when translating into Swedish*

Innehåll

1. Inledning.....	1
2. Om Google Translate	1
2.1 Maskinöversättning	1
2.2 Google Translate	4
3. Metodbeskrivning.....	6
4. Analys och resultat	7
4.1 Allmänt.....	7
4.2 Korrekthet.....	8
4.2.2 Pronomen	8
4.2.3 Artiklar	9
4.2.4 Finita verb och V2-regeln	11
4.2.5 Kongruens	12
4.3 Begriplighet.....	14
4.3.1 Förlorad information	14
4.3.2 Spår av engelska.....	14
4.3.3 Lexikalisk-semantiska problem: Polysemi.....	15
4.3.4 Lexikalisk-semantiska problem: Homonymi	17
4.3.5 Lexikalisk-semantiska problem: Metaforer	18
4.3.6 Prepositioner.....	19
4.4 Variation.....	20
4.4.1 Variation beroende på format (versaler eller gemener).....	21
4.4.2 Variation beroende på kontext	22
4.4.3 Variation vid byte av skiljetecken	24
4.4.4 Variation beroende på källspråk?	25
4.4.5 Variation beroende på målspråk?	25
4.4.6 Sammanfattning av de viktigaste resultaten.....	27

5. Slutdiskussion.....	28
6. Litteraturförteckning	29
Bilagor	32
Bilaga 1: <i>The Economist</i> (2012) och <i>El Economista</i> (2012) omformade till versaler	32
Bilaga 2: Ur Wikipedia: ”Simón Bolívar”	33
Bilaga 3: Kommentarer om googleöversättningar av <i>The Economist</i> (2012) och <i>El Economista</i> (2012)	36

1. Inledning

Med denna uppsats vill jag undersöka maskinöversättningstjänsten Google Translate (GT) och dess möjligheter och begränsningar. Jag vill därtill belysa hur användare av GT kan få varierade resultat genom att alterera källtexters form och kontext.

Mina resultat kommer att delas upp i problemområden som berör korrekthet respektive begriplighet. Problem med korrekthet kan finnas på olika grammatiska nivåer i googleöversättningar. Att kartlägga var dessa problem uppstår kan ge indikationer på var de kan förväntas. Problem som berör begriplighet grundar sig på brister i korrekthet, men de gör sig först gällande när de ändrar eller otydliggör texternas innehåll. Denna uppdelning i korrekthet och begriplighet är relevant eftersom GT inte gör anspråk på att producera korrekta översättningar. En vanlig uppfattning (se avsnitt 2.2.) är också att GT visserligen inte kan prestera korrekta översättningar, men att tjänsten väl kan återge det generella innehållet i t.ex. en nyhetsartikel. Denna uppfattning sätter också begriplighet i ett större fokus än korrekthet.

För att placera GT i sitt sammanhang kommer jag först att ge en bakgrund om tjänsten, dess utveckling och dess tillvägagångssätt. Jag ger också bakgrundsinformation om de mest förekommande typerna av maskinöversättning och redogör för var GT placeras bland dessa. Min metod för att testa GT:s möjligheter och begränsningar kommer sedan att bestå i att närstudera ett kort material av googleöversatt text till svenska från engelska och spanska, och att med denna närstudie som grund laborera med onlinetjänsten för att belysa hur översättningarna kan variera beroende på form och kontext.

Utifrån min undersökning kommer jag att diskutera huruvida GT:s översättningar skiljer sig beroende på genre eller språkpar. De två källspråk som jämförs är engelska och spanska. Målspråket är huvudsakligen svenska, men jämförelse sker också i viss mån med engelska för att avgöra om översättningarna skiljer sig i kvalitet mellan dessa två målspråk.

En förhoppning med denna uppsats är att ge en allmän presentation av GT för icke-insatta. En annan förhoppning är att den ska kunna vägleda översättare och andra användare som ett verktyg för att få ut så mycket som möjligt av tjänsten.

2. Om Google Translate

2.1 Maskinöversättning

Maskinöversättning är automatisk översättning med hjälp av datorer. Till skillnad från andra översättningsprogram som samverkar med människor, s.k. *computer-aided translation tools* (CAT-verktyg), producerar maskinöversättningsverktyg egna texter. Dessa kan användas av översättare för att bearbetas, eller för att utvinna information ur maskinöversättningen som sådan. Intresset för maskinöversättning är nästan lika gammalt som datortekniken själv, och liksom datortekniken har maskinöversättningstekniken utvecklats snabbt (Lopez 2008: 1).

Det finns generellt sett två typer av maskinöversättningsteknik: regelbaserad och korpusbaserad. Regelbaserad maskinöversättning grundar sig på lingvistiska regler, medan den korpusbaserade maskinöversättningen tar hjälp av stora flerspråkliga korpusar för att skapa automatiska översättningar (se t.ex. Ping 2009, 162-163). Wasmuth (2011: 11) beskriver de två typerna av maskinöversättning såhär:

Regelbaserad maskinöversättning är resultatet av ett samarbete mellan språk- och datorexpert. De skapar tillsammans ett mjukvarusystem som används för morfologisk och syntaktisk analys. De sammanställer också ordlistor och grammatiska regler som kan användas av systemet. I korpusbaserad maskinöversättning "lär sig" mjukvaran att översätta genom användningen av stora korpusar som innehåller texter med miljontals sammanlänkade meningar på två språk. De kan liknas vid stora översättningsminnen.

Hartley (2009: 109-110) betonar därtill att de korpusar som används för korpusbaserad maskinöversättning innehåller naturligt språk. Med andra ord är denna variant baserad på text produceras av människor vars beståndsdelar sedan matchas mot källtextens. Renodlade korpusbaserade översättningsverktyg använder sig inte av grammatiska regler alls, utan av bedömningar av ekvivalens på ord- och frasnivå.

Hartley tar också upp svårigheterna med att skapa korpusar för översättning: Parallella, flerspråkliga korpusar kräver så många ord för att fungera i maskinöversättningssyfte att få aktörer kan ha råd att utveckla dem (Hartley 2009: 110).

Trots dessa svårigheter för mindre aktörer dominerar korpusbaserad maskinöversättning idag över regelbaserad. I synnerhet en statistisk variant av korpusbaserad maskinöversättning har varit framgångsrik. Till denna kategori hör bland annat GT. Statistisk maskinöversättning är också den variant som varit föremål för mest forskning (Wasmuth 2011: 12). Den är också den mest kända bland allmänheten, mycket tack vare gratisjänster på internet som just GT.

När Ping (2009) går igenom olika typer av maskinöversättning beskriver han den statistiska varianten såhär:

In **statistical machine translation** (SMT), words and phrases (word sequences) in a bilingual parallel corpus are aligned as the basis for a "translation model" of word-word and phrase-phrase frequencies. Translation involves the selection, for each input word, of the most probable words in the target language, and the determination of the most probable sequence of the selected words on the basis of a monolingual "language model" [...]. Since the translation engine works on the basis of corpora, building quality bilingual text corpora is essential to the success of SMT. (Ping 2009: 163)

Något att ta fasta på i citatet ovan är att statistisk maskinöversättning gör bedömningar på både ord- och frasnivå. De enspråkiga modeller som Ping beskriver involverar regler, men baserat på dessa grundregler gör programvaran jämförelser mellan korpusarna på käll- och målspråket och väljer översättning grundat på statistisk samförekomst.

Lopez (2008: 7-9) tar upp olika modeller för statistisk maskinöversättning. Bland dessa finns både strikt statistiska metoder (ex. "discriminative models", 20-22) och metoder med mer regelbaserade inslag (ex. "generative models", 16-20). Statistisk maskinöversättning är alltså inte begränsad till en och samma modell. Som Ping skriver i citatet ovan är de tvåspråkiga korpusarnas kvalitet av största vikt (jfr. även Hartley 2009: 109-110).

Inom statistisk maskinöversättning har utvecklingen gått i hög fart, och nya idéer introduceras kontinuerligt i tekniken (Lopez 2008: 1). Därför beskriver Lopez teknikens nuvarande status som "a moving target" (2008: 2). Ping (2009: 164-165) nämner också att intresset ökar för röstbaserad maskinöversättning, som översätter ljud istället för text. Detta beskrivs av Ping som en teknik under snabb utveckling.

Ping (2009: 165) tar också upp några generella utmaningar för maskinöversättning. Utmaningarna delas upp i språkliga och icke-språkliga (alltså relaterade till förståelse av omvärlden). Icke-språkliga problem exemplifieras med att människor kan göra bedömningar utifrån en intuitiv slutledningsförmåga som datorer saknar. Det innebär exempelvis att en fråga som "är 35 00 X 58 000 större eller mindre än 1?" av en dator måste besvaras genom en uträkning, medan människor kan sluta sig till ett svar utan att behöva göra någon uträkning (se Ping 2009: 165). Icke-språkliga problem beskrivs som en större utmaning eftersom de är svåra att få en dator att upptäcka.

Ping talar om två huvudtyper av språkliga utmaningar: tvetydigheter (*ambiguity*) och mismatchningar mellan käll- och målspråk.

Tvetydigheterna delas av Ping in i två huvudgrupper, lexikala och grammatisk-strukturella (Ping 2009: 166):

Lexical ambiguity is typically caused by polysemy and homonymy. Structural or grammatical ambiguity arises where different constituent structures (underlying structures) may be assigned to one construction (surface structure).

De grammatisk-strukturella tvetydigheterna delas sedan in i två typfall: skilda strukturer (*alternative structures*) och flertydiga anaforiska referenser (*uncertain anaphoric reference*):

Alternative structures are constructions which present two or more possible interpretations but presuppose that only one is true. For example, from a purely grammatical point of view, "pregnant women and babies" can be interpreted as '(pregnant women) and babies' or 'pregnant (women and babies)', although only the former is semantically accurate.

Uncertain anaphoric reference occurs when an expression can refer back to more than one antecedent, as in the following example: "There's a pile of inflammable trash next to your car. You are going to have to get rid of it." Here, it is not possible to determine, without reference to the context, whether "it" refers anaphorically to "trash" or "car". To disambiguate the second sentence, MT system developers must encode a great deal of real-world knowledge and develop procedures to use such knowledge. (Ping 2009: 166)

Det andra språkliga huvudproblemet enligt Ping, mismatchningar mellan käll- och målspråk, beror på lexikala och strukturella (grammatiska) skillnader mellan språk. På ordnivå skapas ofta problem av att språk kategoriserar ord olika. Ping exemplifierar det med olika kategoriseringar av släktskap: Det engelska ordet "cousin" har exempelvis åtta motsvarigheter på kinesiska (Ping 2009: 167). Ett liknande exempel som involverar svenskan kunde vara att engelskans "grand mother" har två svenska motsvarigheter, som inte är utbytbara. Skilda terminologiska system av detta slag kan relateras till polysemi.

På strukturell nivå uppstår problem när språk använder likadana konstruktioner för olika syften eller olika konstruktioner för samma syfte. Ett exempel är svenskans V2-regel, som innebär att svenskan får en ordföljd som i många fall motsvaras av V3-ordföljd på andra, icke-nordeuropeiska germanska språk.

Wasmuth (2011: 11) hävdar att vissa språkpar är lättare att maskinöversätta mellan än andra. Han menar också att vissa genrer, exempelvis ekonomitexter, i allmänhet är lättare att översätta än andra, exempelvis poesi.

2.2 Google Translate

Ingående förklaringar om hur GT fungerar är svårfunna. Detta kan ha att göra med att Google ser tjänstens utformning som en företagshemlighet som de inte gärna sprider till konkurrenter. Man kan dock hitta beskrivningar av tjänstens metoder i populärvetenskapliga reportage.

Den korpus som GT använder för att ta fram sina översättningar innehåller texter ur många genrer, från FN-dokument till romantext (Kingsley 2011). Annat material kommer från internet, och en del har inkorporerats i samband med Googles bokscanningsprojekt (Adams 2010).

GT:s statistiska tillvägagångssätt innebär att programmet saknar den typ av avancerad lingvistisk kodning som förekommer i regelbaserad maskinöversättning. GT har visserligen en grundläggande språklig modell av hur korrekta meningar ser ut, men utöver detta sker ingen lingvistisk analys av källmaterialet (Kingsley 2011). I stort beskrivs GT:s metoder som framgångsrika och tjänsten har idag en mycket dominerande ställning på marknaden, speciellt i jämförelse med äldre, regelbaserade översättningsprogram. Hartley (2009: 121) anger genomslaget hos GT och andra gratistjänster på internet som en viktig orsak till en växande entusiasm för maskinöversättningens möjligheter. Enligt Adams (2010) är det korpusens storlek som givit GT ett övertag i branschen. Adams menar att tjänsten förbättrades avsevärt mellan 2004 och 2010 på grund av Googles stora tillgång till översatt material att utöka sin korpus med.

När Domeij (2011: 45) redogör för tjänstens historia tar han upp att GT till en början använde sig av ett regelbaserat översättningssystem som kallades *Babelfish*. Den regelbaserade metoden övergavs dock, och Google utvecklade ett eget statistiskt maskinöversättningssystem som först kom i bruk 2006. Detta är inte en utveckling som Google var först med. Adams (2010) beskriver det han menar vara det tidigaste exemplet:

The first significant breakthrough in the potential of mechanised translation came in the early 1990s when IBM produced a model that abandoned any effort to have the computer "understand" what was being fed into it and instead approached the task by installing in the computer the comparative versions of as much translated text as possible and having the system compute the probability of meanings of words and phrases based on statistical precedent. (Adams, 2010)

Detta genombrott var början till en utveckling som sedan fortgick under 90-talet. Enligt Helft (2010) gick forskare på denna tid gradvis över till ett mer statistiskt tillvägagångssätt eftersom de fann att det tillsammans med en tillräckligt stor korpus gav bättre resultat än deras regelbaserade system.

Det finns dock röster som hävdar att effekterna av metoden att ständigt vidga sin korpus för att förbättra kvalitén på översättningarna börjar avta. I Adams (2010) intervjuas till exempel Andreas Zollmann, anställd på GT. Zollmann hävdar att den statistiska metoden har nått vägs ände. Han lägger därför större vikt vid att i framtiden vidareutveckla de regelbaserade metoderna.

GT utvecklas fortfarande kontinuerligt. Helft (2010) beskriver till exempel planer från Googles sida på att kombinera tjänsten med en scanfunktion som skulle möjliggöra att fotografera text i sin fysiska omgivning för omedelbar maskinöversättning.

Enligt Kingsley (2011) tar GT inte hänsyn till språkriktighet, utan bara till den faktiska användningen av språket. Han menar att tjänsten därför är bra på att översätta ledigt språk. En farhåga som han uttrycker är att Google i framtiden kan komma att få monopol på språklig korrekthet, grammatiska regler etc. Kingsley befarar också att tjänster som GT till slut kan bli så bra att många köpare av översättningar nöjer sig med dem istället för mänskliga översättare för att det är billigare (vilket redan sker i vissa fall). Slutligen ser Kingsley en risk att människors intresse för att själva lära sig språk kan minska i framtiden. Helft (2010) delar dock inte Kingsleys analys:

Automated translation systems are far from perfect, and even Google's will not put human translators out of a job anytime soon. Experts say it is exceedingly difficult for a computer to break a sentence into parts, then translate and reassemble them. (Helft, 2010)

Kingsley (2011) nämner också en begränsning hos programmet, nämligen att det (kanske i likhet med all maskinöversättning?) inte har tillgång till den kulturella kontexten och därför inte kan göra samma val som en människa (jfr. Pings resonemang i 2.1, s. 3-4). Andreas Zollmann tar också upp exempel på grammatiska översättningsproblem när han intervjuas av Adams: "Pronouns, say, are very difficult in some languages where the masculine and feminine don't correspond to each other." (Adams, 2010). Zollmann hävdar samtidigt att ingen forskare förväntar sig att resultaten någonsin kommer att bli perfekta.

En vanlig uppfattning är att GT är tillräckligt bra för att återge texters innehåll på makronivå, exempelvis kärninnehållet i en nyhetsartikel (Helft 2010). GT används enligt Helft av miljoner människor men genererar trots detta inte några stora pengar i jämförelse med exempelvis Googles reklamintäkter.

Shen (2010) hävdar i en studie att översättning med GT fungerar speciellt bra mellan engelska och en handfull EU-språk (framför allt franska, men också italienska, tyska och

portugisiska). Shen inkluderar här spanska som målspråk men inte som källspråk med engelska som målspråk. Han anger som möjlig orsak att många latinamerikanska länder erbjuder engelska översättningar av centrala offentliga dokument. Generellt pekar Shens studie på att engelska generellt ger godare resultat som målspråk än genomsnittet.

3. Metodbeskrivning

För att belysa GT:s möjligheter och begränsningar har jag för det första valt att detaljgranska två korta utdrag (ca 200 ord var) ur ekonomiartiklar, en på engelska och en på spanska (*The Economist* 2012 och *El Economista* 2012, båda osignerade) för att få en generell bild av vilka språkliga problem och möjligheter man kan träffa på i googleöversättningar. Jag problematiserar sedan resultaten från denna undersökning genom att på ett friare och mer experimenterande sätt laborera med de fenomen jag finner vanliga i tjänstens översättningar. Därigenom vill jag också belysa hur beroende programvaran är av form och kontext.

Som nämnts i min inledning delar jag upp mina resultat i problemområden som berör korrekthet och sådana som avser begriplighet. Detta val kan relateras till det resonemang som förts av Kingsley (2011), Adams (2010) och Helft (2010) refererade i 2.2, och som utgår från att GT aldrig kan förväntas översätta helt korrekt, men dock kan användas för att tillgodogöra sig texters innehåll på en mer generell nivå.

För min metod används inte den traditionella uppdelningen i språkliga nivåer i grammatiken (morfologi, lexiologi och syntax). Detta val kan relateras till Pings beskrivning av statistisk maskinöversättning som en metod som behandlar såväl ord- som frasnivå samtidigt (se avsnitt 2.1). Det kan också relateras till Kingsleys (2011) påpekande att GT inte tar hänsyn till språkriktighet (se avsnitt 2.2.). I GT:s översättningar påverkas det syntaktiska av det lexikala och morfologiska och vice versa. Tyngdpunkten faller därför varken på ord, fraser eller meningar, utan på allt samverkande. Snarare än att försöka kategorisera problemen i språkliga nivåer (ytterligheter) vill jag således beskriva vilka uttryck dessa problem kan ta sig.

Av utrymmesskäl kommer min studie inte att ta upp alla de fenomen jag noterat. Jag kommer istället att fokusera på ett antal generella problemområden som är särskilt återkommande.

En särskild svårighet vid analys av GT:s översättningar är att GT, i likhet med statistisk maskinöversättning i allmänhet, är i ständig utveckling (se t.ex. Lopez 2008: 2, citerad i 2.1).

Detta medför att de översättningar som produceras av tjänsten idag med största sannolikhet inte blir desamma som de översättningar tjänsten producerar vid ett senare tillfälle. Med andra ord är de empiriska studier som ligger till grund för mina resultat troligen omöjliga att upprepa (såväl av mig själv som av någon annan) när denna uppsats publiceras, och framtida resultat riskerar att bli mer annorlunda gentemot mina ju längre tiden går. Min metod kan heller inte väntas ge en diakron beskrivning av GT:s utveckling.

Att dra slutsatser om orsaker till vanliga grammatiska fel i GT:s översättningar är svårt eftersom varken tjänstens hela korpus eller någon ingående beskrivning av tjänstens teknik finns tillgängliga. Dock finns Googles vanliga sökmotor tillgänglig för allmänheten, och en del val som GT gör kan relateras till de grammatiska alternativ som är vanligast förekommande i Googles sökmotor. Om en viss grammatisk konstruktion får färre träffar än en annan möjlig konstruktion i Googles sökmotor borde detta kunna vara en indikation på att den också kommer att vara mindre förekommande i GT:s översättningar. Eftersom Googles sökmotor innefattar mycket mer material än GT:s korpus kan vi inte få något exakt svar utifrån den, men den kan ändå ge oss ledtrådar om hur GT:s statistiska approach påverkar översättningarna.

4. Analys och resultat

Jag har valt att dela upp mina resultat i underavdelningarna korrekthet (4.2) och begriplighet (4.3). Inom båda dessa underavdelningar tar jag också upp hur översättningarna kan variera vid alterering av källtexten. En djupare analys av dessa variationer sker sedan i avsnitt 4.4. Allra först görs några allmänna påpekanden om GT.

4.1 Allmänt

Utan tvekan har GT många starka sidor, trots de problem jag kommer att ta upp i senare avsnitt. Till GT:s uppenbara fördelar (och kanske det som gjort tjänsten så populär) hör att tjänsten är snabb och kostnadsfri: Användaren kan på en sekund och helt gratis få tillgång till föreslagna översättningar som hade varit både dyrare och mer tidskrävande om de producerats av mänskliga översättare.

Kvalitetsmässigt klarar GT i många situationer av att prestera korrekta översättningar av konstruktioner, samtidigt som samma konstruktioner i andra kontexter vållar problem. På

övergripande meningsnivå finns dock få exempel på helt korrekta översättningar. I mitt källmaterial (*The Economist* 2012 och *El Economista* 2012) finns bara ett exempel:

- (1) **KT:** Judging by the numbers, the RBI is among the world's best central banks.
GT: Att döma av siffrorna är RBI bland världens bästa centralbanker. (5/2 2012)

Men det är också möjligt att framhålla kvalitéer i meningar som inte är fullt korrekta, som följande:

- (2) a. **KT:** ONE of the perks of *being* governor of the Reserve Bank of India...
GT: EN av de förmåner med att vara guvernör Reserve Bank of India...(5/2 2012)
- b. **KT:** Berlín estudia subvencionar a Grecia a través del *consumo eléctrico alemán*
GT: Berlin studerar att subventionera Grekland genom den tyska elförbrukningen (5/2 2012)
- c. **KT:** El semanario basa sus informaciones en **planes de la Comisión de la Unión Europea** para un nuevo programa *de fomento del crecimiento económico* en Grecia.
GT: Den veckovisa bygger sin information om planer på EU-kommissionen för ett nytt program för att främja den ekonomiska tillväxten i Grekland. (5/2 2012)

I exempel (2a) klarar GT av att översätta en engelsk ing-form korrekt (till en svensk infinitivfras). I (2b) presterar den en fungerande sammansättning på svenska och avgör korrekt vilket av de två spanska adjektiven som blir ett led i en svensk sammansättning respektive ett adjektivattribut. I (2c) klarar GT av att göra om en nominalfras till en korrekt svensk infinitivfras.

Som vi ska se är dessa exempel dessvärre inte alltid representativa.

4.2 Korrekthet

Ser man på de översättningar jag använder som huvudsakligt källmaterial finns vid första anblick en rad brister. Vad man sedan kan se är dock, som jag varit inne på, att ett grammatiskt drag kan översättas inkorrekt i ett sammanhang för att i ett annat bli korrekt. Detta kan ha att göra med tjänstens beroende av kontext, vilken jag kommer att återkomma till i avsnitt 4.4. Nedan beskrivs några huvudområden där tjänsten ofta får problem med korrekthet.

4.2.2 Pronomen

- (3) **KT:** India's central bank is one of *its* best institutions.

GT: Indiens centralbank är ett av *sina* bästa institutioner (5/2 2012)

I exempel (3) översätts engelskans possessiva pronomen *its* till *sina*. *Its* motsvarar på svenska både ett reflexivt pronomen (*sin/sitt/sina*) och possessivt (*dess*). Liknande exempel på fel pronomen går att hitta, även från spansk källtext, som här:

- (4) **KT:** El semanario basa sus informaciones en **planes de la Comisión de la Unión Europea** para un nuevo programa de fomento del crecimiento económico en Grecia. Agrega que, sin embargo...

GT: Den veckovisa bygger sin information om planer på EU-kommissionen för ett nytt program för att främja den ekonomiska tillväxten i Grekland.
Han tillägger dock att... (5/2 2012)

När *Den veckovisa* (alltså tidskriften *Der Spiegel*) ska ges en identisk anafor i följande mening väljer GT pronomenet *han* och tolkar därmed tidsskriften som en (maskulin) person. Formaterar man om samma källmening till endast versaler (se bilaga 1) står subjektet inte med alls i översättningen: ”Tillagt att dock ...”. (Mer om dylika omformateringar i avsnitt 4.4.)

Man kan spekulera i om ett sådant här fel kan härledas till spanskans avsaknad av subjektstvång (och avsaknaden av subjektspronomen i den andra källmeningen i detta exempel). Dock kan vi, som med andra fenomen, inte veta säkert eftersom grammatiska regler inte spelar avgörande roll för maskinvaran. Vi kan dock konstatera att (3) innehåller ett engelskt pronomen med flera möjliga ekvivalenter på svenska beroende på kontext. Således uppstår här ett problem som kan relateras till de båda språkens olika pronomensystem, där svenskan har en distinktion mellan reflexiv/icke-reflexiv syftning (jfr. Ping 2009, refererad i avsnitt 2.1).

4.2.3 Artiklar

- (5) **KT:** ONE of the perks of being governor of the Reserve Bank of India (RBI) is the use of a colonial bungalow on Carmichael Road...

GT: EN av *de* förmåner med att vara guvernör Reserve Bank of India (RBI) är användningen av en kolonial bungalow på Carmichael Road...(5/2 2012)

Den svenska konstruktionen i (5) hade varit korrekt om den följs av en relativ bisats: *de förmåner som finns med att....* I denna målspråksmening vore det dock mer troget källtexten med enbart morfologisk bestämning, *förmånerna med att....*

Man kan ana en möjlig orsak till just detta fel om man undersöker förekomsten av de båda alternativen i googles sökmotor; ”de förmåner med” ger 18 200 träffar, medan en på ”förmåner med” bara ger 11 600 (3/5 2012).

Det finns emellertid andra passager där tjänsten klarar av svenskans artikelsyntax, exempelvis dubbel bestämning:

- (6) **KT:** Feb 4th 2012 | MUMBAI | *from the print edition*
GT: 4 feb 2012 | Mumbai | *från den tryckta upplagan* (5/2 2012)

Man kan notera att Googles sökstatistik här talar till det korrekta alternativets fördel. Frasen ”den tryckta upplagan” ger 56 100 träffar på Google medan ”den tryckta upplaga” bara ger 8 (3/5 2012). Av dessa tycks två vara inkorrekt använda, medan konstruktioonen i övriga exempel efterföljs av en bisats, t.ex. ”den tryckta upplaga som...”.

I min exempelöversättning på ca 200 ord från spanska klarar tjänsten av att anpassa många av artiklarna till svenskan korrekt, som här:

- (7) **KT:** El Gobierno de la canciller federal, Angela Merkel, estudia la posibilidad de subvencionar *la debilitada* economía de Grecia, sumida *en la crisis* de la deuda, *a través del consumo* de electricidad en Alemania, según el semanario *Der Spiegel*.

GT: Regeringen i förbundskanslern är Angela Merkel överväger att subventionera *den krisdrabbade ekonomin* i Grekland, fast *i skuldkris*, *genom konsumtion* av el i Tyskland, enligt veckans *Der Spiegel*. (5/2 2012)

Här förmår tjänsten översätta en bestämd konstruktion till en obestämd och därmed anpassa den till svenskt språkbruk (spanskan brukar bestämd artikel mer än svenskan). Denna skillnad blir svår att förklara med engelskans och spanskans olika artikelsystem eftersom de i detta fall är påfallande lika (båda språken saknar morfologisk artikel). Dock kan återigen konstateras att detta falls korrekta kollokation, ”genom konsumtion av” får 14 700 träffar på Googles sökmotor mot 776 för ”genom konsumtionen av” (3/5 2012).

Vi ser också att GT väljer att sätta ut dubbel bestämning i konstruktionen *den krisdrabbade ekonomin*. Detta val kan också stödjas med en jämförelse mellan sökningar på ”den krisdrabbade ekonomin” (17 000 träffar) med ”den krisdrabbade ekonomi” (3 träffar, 3/5 2012). Vi ser således ytterligare samband mellan tjänstens val och den statistiska förekomsten av valda kollokationer i Googles sökmotor.

Sambandet mellan GT:s val och det givna alternativets förekomst i Googles sökmotor är dock inte självklart. Frasen ”i skuldkrisen” får exempelvis 21 200 träffar mot tjänstens här

valda ”i skuldkris” som får 2 830 (3/5 2012). Man måste därmed sluta sig till att också andra faktorer har inverkan. (Se vidare avsnitt 4.4 nedan).

4.2.4 Finita verb och V2-regeln

Både översättningar från engelska och från spanska till svenska med GT innehåller extrainsatta finita verb som gör meningarna ogrammatiska. I min granskade googleöversättning av det spanska utdraget på ca 200 ord rör det sig om kopulaverb, medan det engelska textutdraget har dubbleringar av verbet *har* och ett extrainsatt *kan*.

Dubbleringen av verbet *har* i översättningen från engelska kan ha att göra med den svenska V2-regeln; verbet *har* dubbleras exempelvis i mening nedan, där engelskan har V3-ordföljd:

- (8) **KT:** Since 1995 wholesale prices *have* risen by an average of a year, not too far from the RBI's comfort zone of about 5%.

GT: Sedan 1995 *har* grossistpriserna *har* stigit med i genomsnitt 6% per år, inte alltför långt från Indiens centralbank komfort zon ca 5%. (5/2 2012)

Att ha kvar verbet på motsvarande plats i källtexten i denna exempelmening är ogrammatiskt på svenska. Felet är med andra ord inte att tjänsten sätter in ett verb på sin naturliga plats i det svenska satsschemat, utan att verbet samtidigt står kvar på den engelska ursprungsplatsen.

Denna exempelmening översätts på samma sätt även om den isoleras från resten av texten.

Dock kan man också ta fram exempel där tjänsten klarar av att placera om verbet i samma typ av konstruktion:

- (9) **KT:** Since then, I have tried to become a member of the team.

GT: Sedan dess har jag försökt att bli medlem i teamet. (27/2 2012)

Här kan vi, liksom i föregående avsnitt, konstatera att statistiken från Googles sökmotor talar för det korrekta alternativet. ”Sedan dess har jag” ger 2 460 000 träffar mot 118 000 för ”Sedan dess jag har”. De flesta av dessa senare exempel är också av typen *Jag har varit aktiv ända sedan dess. Jag har bland annat...*

I mitt källmaterial från spanska finns en mening med en ordföljd som mycket väl passar på svenska, men som ändå fått ett tillagt kopulaverb:

- (10) a. **KT:** El Gobierno de la canciller federal, Angela Merkel, estudia la posibilidad de subvencionar la debilitada economía de Grecia, sumida en la crisis de la deuda, a través del consumo de electricidad en Alemania, según el semanario Der Spiegel.

GT: Regeringen i förbundskanslern *är* Angela Merkel överväger att subventionera den krisdrabbade ekonomin i Grekland, fast i skuldkris, genom konsumtion av el i Tyskland, enligt veckans Der Spiegel. (5/2 2012)

Här hade det som sagt varit korrekt att bara ha kvar verbet från källtexten på motsvarande position (*Förbundskansler Angela Merkel överväger...*). Det är också denna översättning vi får om vi tar bort Angela Merkels titel, eller exempelvis lägger till en annan:

- (10) b. **KT:** Angela Merkel estudia la posibilidad de subvencionar la debilitada economía de Grecia, sumida en la crisis de la deuda, a través del consumo de electricidad en Alemania, según el semanario Der Spiegel.

GT: Angela Merkel överväger att subventionera krisdrabbade ekonomin i Grekland, fast i skuldkris, genom konsumtion av el i Tyskland, enligt veckans Der Spiegel. (26/4 2012)

- c. **KT:** La química física Angela Merkel estudia la posibilidad de subvencionar la debilitada economía de Grecia, sumida en la crisis de la deuda, a través del consumo de electricidad en Alemania, según el semanario Der Spiegel.

GT: Den fysikalisk kemi Angela Merkel överväger att subventionera krisdrabbade ekonomin i Grekland, fast i skuldkris, genom konsumtion av el i Tyskland, enligt veckans Der Spiegel. (26/4 2012)

Även om tjänsten inte klarar av att översätta Angela Merkels alternativa titel (forskare i kvantkemi) finns inget extrainsatt kopulaverb i den svenska översättningen av (10c). Inte heller i (10b), där Angela Merkels titel tagits bort. Detta belyser ånyo tjänstens beroende av kontext (där även borttagen text kan inkluderas som ett sätt att underlätta för GT att producera mer korrekta översättningar, kanske speciellt när man har att göra med längre, mer komplexa meningar).

4.2.5 Kongruens

- (11) **KT:** India's central bank is one of its best institutions. *It is also complicit* in a government-borrowing binge

GT: Indiens centralbank är ett av sina bästa institutioner. *Det är delaktig* även i ett statligt lån hetsätning (5/2 2012)

Vi ser här ett typiskt kongruensproblem där GT böjer efter fel genus på svenska. Det finns dock också tillfällen där kongruensen fungerar bra:

- (12) **KT:** Judging by the numbers, the RBI is among *the world's best* central banks. Its record on balancing growth and inflation is decent enough (see chart 1).

GT: Att döma av siffrorna är RBI bland *världens bästa* centralbanker. Dess resultat när det gäller balansering tillväxt och inflation är *anständiga* nog (se diagram 1). (5/2 2012)

Felet i exempel (11) kan inte förklaras av förekomsten på Googles sökmotor eftersom det korrekta ”den är delaktig” får 214 000 träffar mot ”det är delaktig” med 5 (3/5 2012). Sökningen på ”världens bästa” får samtidigt 12 700 000 träffar på Google mot ”världens bäst” som får 66 100 (där vanliga exempel är av typen *världens bäst säljande/rankade...*).

Bortsett från exemplet ”världens bästa” kan kongruensexemplen (11) och (12) ovan inte sägas innehålla särledes konventionella standardfraser. Testar man tjänsten med mer konventionaliserade engelska fraser och satser tycks dock chanserna för korrekta kongruensformer öka:

- (13) a. **KT:** Roses are red
GT: Rosor är röda (27/4 2012)
- b. **KT:** The Swedish government
GT: Den svenska regeringen (27/4 2012)
- c. **KT:** The invincible armada
GT: Den oöverbärliga armada (27/4 2012)
- d. **KT:** Alexander the great
GT: Alexander den store (27/4 2012)

Kollokation och konventionalisering tycks alltså vara en viktig faktor vid GT:s hantering av svenskans adjektivkongruens. Ytterligare en faktor skulle kunna vara de konventionaliserade uttryckens kontext. Detta stöds av att korrektheten hos översättningarna av dessa kollokationer kan försvinna så fort kollokationen delas upp och någon annan information hamnar emellan:

- (14) **KT:** Roses are, in general, red
GT: Rosor är i allmänhet, röd (27/4 2012)

En standardiserad kollokation tycks med andra ord bara uppfattas som sådan om den står för sig själv.

Denna typ av kongruensproblem hittar jag inte i min exempelöversättning på 200 ord från spanska. Det skulle möjligen kunna relateras till att spanskan kongruensböjer adjektiv efter numerus och genus precis som svenskan, men för att testa ett sådant påstående på allvar skulle ett större material behöva studeras.

De grammatiska, morfologiska och syntaktiska felen i GT:s översättningar är som nämnts i avsnitt 3 inte oväsentliga vare sig för översättningarnas användningsmöjligheter eller för hur de uppfattas. De torde däremot i de flesta fall ha liten inverkan på texters begriplighet.

4.3 Begriplighet

Eftersom GT inte gör anspråk på att kunna producera helt korrekta översättningar är problem med korrekthet väntade (och vissa av dem också omnämnda av Kingsley m.fl.). Det område som dock mer berör programvarans tilltänkta användning, såsom t.ex. Kingsley beskriver den i 2.3, är begriplighet. Nedan kommer jag att ta upp fenomen som på ett mera avgörande sätt kan tänkas påverka måltexternas begriplighet och möjligheten att tillgodogöra sig innehållet.

4.3.1 Förlorad information

Det händer att ord helt försvinner vid översättning med GT. I vissa meningar kan ord som är avgörande för meningens grammatiska korrekthet och förståelsen av innehållet försvinna:

- (15) **KT:** Para incentivar las inversiones en plantas solares o parques eólicos, las subvenciones públicas alemanas a las energías renovables podrían beneficiar también a empresas griegas que suministren *electricidad* a la red europea, añade el semanario.

GT: För att uppmuntra investeringar i solenergi växter eller vindkraftverk, de tyska offentliga subventioner till förnybar energi skulle också kunna gynna grekiska företag som levererar [ord saknas, förf. anm.] till det europeiska nätverket, tillade vecka. (5/2 2012)

Om man har textens sammanhang kanske man kan även här kan förstå att det är el som de grekiska företagen levererar, men utelämnningen försvårar ändå läsbarheten. Sådana här utelämnningar av ord kan få mer eller mindre allvarliga konsekvenser beroende på om ordet i fråga är centralt för förståelsen av innehållet eller inte.

4.3.2 Spår av engelska

I mitt översättningsmaterial från engelska förekommer ord från källtexten som lämnats oöversatta. De har alltså inte utelämnats, utan är kvar i sin ursprungliga, engelska form:

- (16) a. **KT:** Pulling every lever
GT: *Pulling* varje spak (5/2 2012)

- b. **KT:** Carmichael Road, a posh street
GT: Carmichael Road, en *posh* gata (5/2 2012)

Förståelsen hos en svensk beror här delvis på hur goda engelskkunskaper en användare har, men även om läsaren skulle ha goda engelskkunskaper störs läsningen och kanske också begripligheten. Om användaren översätter från ett annat källspråk än engelska kan såväl ord på såväl originalspråket i fråga som på engelska förekomma i översättningen.

Vid översättningen av min spanska källtext (*El Economista*, 2010) förekommer endast engelska ord om den översätts formaterad till endast versaler. (Jag återkommer till detta i avsnitt 4.4.1 nedan).

4.3.3 Lexikalisk-semantiska problem: Polysemi

Polysemi är ett centralt problem för GT:s förmåga att skapa begripliga översättningar. Det har att göra med att polysemi innebär flera möjliga alternativ, som GT måste välja mellan utifrån bedömningar om statistisk samförekomst. Polyseminproblem finns genomgående i GT:s översättningar. Nedan tre exempel där problematiken visar sig på lexikal nivå:

- (17) a. **KT:** ONE of the perks of being *governor* of the Reserve Bank of India...
GT: EN av de förmåner med att vara *guvernör* Reserve Bank of India... (5/2 2012)
- b. **KT:** the RBI's *comfort zone*
GT: Indiens centralbank *komfort zon* (5/2 2012)
- c. **KT:** Para incentivar las inversiones en *plantas* solares o parques eólicos...
GT: För att uppmuntra investeringar i solenergi *växter* eller vindkraftverk... (5/2 2012)

Sådana här polyseminproblem kan leda till större eller mindre problem för begripligheten. Exemplet ”komfort zon” i (17b) torde inte sänka förståelsen för de flesta. I (17c) är det dock lätt att tro att det handlar om investeringar i ”växter”, när det i själva verket handlar om solceller. Det är dock möjligt att laborera med ordet ”planta” för att betydelsen ’kraftverk’ ska komma fram. Låt oss se på hur översättningen av den spanska kollokationen ’Plantas solares’ ändras allteftersom vi skriver in den i GT:

(18)

KT	GT 13/4 2012
plantas s	fabriker

plantas so	växter så
plantas sol	sol växter
plantas sola	enstaka anläggningar
plantas solar	solkraftverk
plantas solare	Solare växter
plantas solares	sol
plantas solares.	solanläggningar

Tolkningen av det polysema ordet *plantas* växlar som vi ser konstant. Avgörande för hur ordet översätts är dessutom om vi sätter ut en punkt efter ordet eller inte. Att bruket vacklar i början av inskrivningen är inte förvånande, eftersom inte heller en människa helt kan utröna ordets betydelse. Vid de tre sista bokstäverna finns dock inte längre mycket möjlighet för en människa till andra tolkningar än att det rör sig om kraftverk.

Skriver vi in andra sorters kraftverk på spanska, med mer entydigt elektricitetsrelaterade adjektiv, blir det dock lättare att få ut kollokationernas betydelse direkt och även utan att sätta någon punkt.

- (19) a. **KT:** planta eléctrica
GT: kraftverket (13/4 2012)
- b. **KT:** planta nuclear
GT: kärnkraftverk (13/4 2012)

Tar vi slutligen med en av dessa kollokationer tillsammans med attributet *solares* klarar GT av att översätta också detta komplexa uttryck, och någon punkt behövs inte):

- (19) c. **KT:** plantas nucleares y solares
GT: nukleära och solkraftverk (13/4 2012)

De variationer jag tagit upp så här långt illustrerar att polysemi är en central utmaning för GT, liksom för övrigt för statistisk maskinöversättning i allmänhet. Ett ords olika översättningsalternativ är beroende av ordets kontext. Här omtolkas kontext som statistisk samförekomst. Detta innebär att en mer oväntad ordkonstellation i källtexten troligen löper större risk att felöversättas (jfr. mitt resonemang om kongruens vid mer eller mindre konventionella fraser i 4.2.5 ovan.)

På grund av den utmaning polysemi innebär är det möjligt att sluta sig till att GT i allmänhet får svårare att översätta begripligt ju mer frekvent polysemi är i en text. Detta kan illustreras om man låter tjänsten översätta en ordlek byggd på polysemi:

- (20) **KT:** If you keep to the left you are sure to go right, if you go to the right you go wrong.
GT: Om du håller till vänster är du säker på att gå till höger, om du går till höger du går fel. (13/4 2012)

Förekomsten i KT av *right* med två olika betydelser blir här för mycket för GT, särskilt som de bägge orden uppträder i en konstellation (*go right*) som kan ges bägge tolkningarna. Skriver man in ett fragment av meningen kan man också i detta fall se att GT översätter ordet olika beroende på om man väljer att sätta en punkt eller inte:

- (21) a. **KT:** You are sure to go right
GT: Du är säker på att gå rätt (13/4 2012)
- b. **KT:** You are sure to go right.
GT: Du är säker på att gå till höger. (13/4 2012)

För en mänsklig språkvarelse kan denna mening, med eller utan punkt, sällan eller aldrig fungera med betydelsen 'höger'. Intressant är också att tjänsten vid översättning av (21a) föreslår både 'höger' och 'direkt' som alternativ till 'rätt', medan den inte föreslår några alternativ alls till 'höger' i (21b). (Mer om punkter och andra skiljetecken i 4.4.2 – 4.4.3 nedan.)

De resultat jag nu presenterat om polysemi ger för det första en bild av den utmaning detta område innebär för GT. De antyder också att man kan få fram en större begriplighet i GT:s översättningar genom att laborera med dem, om man vet hur man ska göra.

4.3.4 Lexikalisk-semantiska problem: Homonymi

I mitt huvudkällmaterial saknas exempel på problem relaterade till homonymi. Detta kan ha att göra med att de texter jag valt att närstudera är relativt korta. Studerar man andra exempelfraser översatta av GT är det inte svårt att hitta problem som skapats av homonymi. Nedan ett exempel på en sådan översättning av en engelsk konstruktion med ett ord som är homonym till ett annat ord. Notera hur tjänsten väljer olika alternativ beroende på kontext:

- (22) a. **KT:** They wanted to excel
GT: De ville till Excel (25/4 2012)

- b. **KT:** They wanted to excel in
GT: De ville utmärka sig i (25/4 2012)
- c. **KT:** They wanted to excel in their studies.
GT: De ville att utmärka sig i sina studier. (25/4 2012)

Detta exempel har mycket gemensamt med de polysemirelaterade exempel ovan av naturliga skäl: I grunden rör det sig om samma problem för tjänsten, att det finns mer än en betydelse att välja mellan. Precis som med polysemi kan detta också leda till problem med källtexters begriplighet när tjänsten väljer fel, som i (22a), där man kan missta innehållet om att 'utmärka sig' för att istället vilja ha ett datorprogram.

4.3.5 Lexikalisk-semantiska problem: Metaforer

Andra oväntade betydelser hos ord och uttryck som vållar problem vid översättning återfinns i metaforer (och säkerligen också i metonymi, vilket dock inte finns exemplifierat i mitt huvudkällmaterial). Metaforer som utmaning för GT liknar polysemi, vilket kan ha att göra med att polysemi i många fall bygger på just metaforer.

Ett exempel på när metaforer har orsakat problem för GT finns i exempel (11) ovan, där det metaforiskt använda *borrowing binge* översätts bokstavligt till *lån hetsätning*. Egentligen beskriver denna formulering i källtexten de ohämmade eller omåttliga låneexcesser som Indiens centralbank är delaktig i. I den anglosaxiska världens ekonomiartiklar är det inte ovanligt att uttrycka sig metaforiskt, ibland på ett drastiskt sätt liknande detta exempel. Det blir dock uppenbart att genrekonventioner på en så pass detaljerad nivå svårligen känns igen av GT. Bokstavliga tolkningar av metaforer av detta slag riskerar att påverka begripligheten negativt.

Som i många andra fall förekommer dock även här variation vad gäller utfallet. Vissa metaforer översätts korrekt, och i vissa fall väljer tjänsten aktivt en annan metafor som också passar tillfället:

- (23) **KT:** Agrega que, sin embargo, **umentan entre tanto las críticas hacia Atenas** por su dubitativa actitud a la hora de afrontar la crisis financiera que *sacude al país*.
GT: Han tillägger dock att öka kritik från både Aten för sin tveksamma inställning i hanteringen av finanskrisen *sveper över landet*. (5/2 2012)

I källtexten talas det här om den finanskris som *skakar* landet, medan den i översättningen *sveper över* landet. Även om metaforerna inte är identiska, och det egentligen vore mer korrekt (i bemärkelsen troget källtexten) att använda den ursprungliga metaforen här, är denna andra metafor funktionell och troligen lika bidragande till översättningens begriplighet.

Även metaforer kan testas på samma sätt som kongruens och polysemi. Standardiserade levande metaforer (se t.ex. Lundquist 2005) hittar ofta sina motsvarigheter om de står för sig själva:

- (24) a. **KT:** Los políticos reafirman la necesidad de *apretarse el cinturón*.
GT: Politiker bekräftar behovet av att *dra åt svångremmen*. (27/4 2012)
- b. **KT:** Politicians reaffirm our need to *tighten our belts*.
GT: Politiker bekräftar vårt behov av att *dra åt svångremmen*. (27/4 2012)
- c. **KT:** Politicians reaffirm the need to *tighten the economy*.
GT: Politiker bekräftar behovet av att *strama åt ekonomin*. (27/4 2012)

Stelnade eller döda metaforer torde generellt vara enklare för GT att översätta eftersom de till sin natur är standardiserade. Men även standardiserade levande metaforer som *dra åt svångremmen* kan alltså fungera bra när de står för sig själva. Mer kreativa metaforer som *borrowing binge* i exempel (11) blir ett problem för GT av naturliga själv, eftersom det är mer osannolikt att de skulle finnas representerade i en korpus.

4.3.6 Prepositioner

Prepositionerna tycks vålla problem antingen genom att vara felöversatta eller genom att inte vara översatta alls, vilket är avgörande för om vissa meningar blir grammatiska eller inte (och ibland inverkar på texters begriplighet). Nedan ett exempel från engelska med tre prepositionsfel:

- (25) **KT:** On one side live some of India's richest industrialists, modern-day pharaohs with flashy architectural tastes. *On the other*, a stone's throw down a cliff, is a small slum — a monument to desperation and government failure.
- GT:** På ena sidan bor några av Indiens rikaste industrimän, moderna faraoner med flashiga arkitektoniska smak. Å *andra sidan* ett stenkast ner en klippa, är en liten slum - ett monument [preposition saknas, förf. anm.] desperation och regeringen misslyckande. (5/2 2012)

Om vi börjar med kollokationerna *on one side – on the other side*, är de båda rumsadverbial på engelska, medan den sista av dem översätts med ett talaktsadverbial, *å andra sidan*, på

svenska. *On the other side* kan dock aldrig ha en sådan funktion på engelska (*On the other hand* hade haft det).

I detta exempel ser vi också prepositionen *down*, som översatts *ner*. *Down* kan fungera både som ett riktningsadverbial och som en preposition. Svenskans *ner* kan dock aldrig fungera som preposition.

Slutligen utelämnas en svensk översättning av prepositionen *to* på rad 2, vilket gör denna exempelmening ogrammatisk. De två första kollokationerna *on one side* och *with flashy architectural taste* översatts dock med korrekta svenska prepositioner.

Spanska som källspråk ger liknande problem. Vissa av dessa har dock sin grund i att spanskan bara konstruerar genitiv med preposition. Nedan två exempel:

- (26) a. **KT:** El Gobierno *de* la canciller federal, Angela Merkel...
GT: Regeringen *i* förbundskanslern är Angela Merkel... (5/2 2012)
- b. **KT:** El semanario basa sus informaciones *en planes de la Comisión de la Unión Europea* para un nuevo programa *de* fomento del crecimiento económico en Grecia.
GT: Den veckovisa bygger sin information om planer på EU-kommissionen för ett nytt program för att främja den ekonomiska tillväxten i Grekland. (5/2 2012)

Både (26a) och (26b) är exempel på spansk genitiv vars prepositioner misstolkas, vilket gör de svenska meningarna svårtolkade. Däremot översatts egennamnet *EU-kommissionen* (som är en genitivkonstruktion på spanska) korrekt. Det vanliga tycks dock vara att GT inte förmår översätta spanska prepositionsfraser med exempelvis sammansättningar, vilket annars är en vanlig och ofta korrekt lösning.

Prepositionsproblemen tar sig liknande uttryck både när källspråket är engelska och när det är spanska. Problemen kan relateras till polysemi på så vis att en preposition på ett språk kan motsvaras av flera alternativa prepositioner (eller en helt annan typ av konstruktion) på målspråket, och vice versa.

4.4 Variation

De studier jag har gjort av GT:s översättningar visar att de kan variera beroende på format och kontext. Man kan således laborera med tjänsten på olika sätt för att få fram annorlunda resultat. Nedan presenteras några altereringar av textmaterial som inte sällan inverkar på hur översättningarna utfaller.

4.4.1 Variation beroende på format (versaler eller gemener)

Översättningar genom GT blir ofta annorlunda om man formar om texterna till enbart versaler. Vissa fenomen som hittas vid översättning från källtexten i vanligt format försvinner vid översättning från versaler. Andra tillkommer. Låt oss ta ett exempel (bokstaven V framför en av översättningarna står för 'versaler').

(27) **KT:** India's central bank is one of its best institutions. It is also complicit in a government-borrowing binge

GT: Indiens centralbank är ett av sina bästa institutioner. Det är delaktig även i ett statligt lån hetsätning (5/2 2012)

GT (V): INDIEN centralbank är en av de bästa institutionerna. DET medverkar även i en regering som lånar SUPKALAS (26/3 2012)

Olikheterna i dessa båda översättningar är flera: Versalöversättningen återger inte genitivformen, tolkar det possessiva pronomenet som en bestämd artikel, tolkar sammansättningen *government borrowing* som *en regering som lånar* och tolkar metaforen i slutet av meningen annorlunda (dock också bokstavligt).

Förutom denna variation finns det enligt mina resultat också en skillnad i vilka engelska ord som lämnas kvar oöversatta i måltexten. Exempelmeningen *Pulling every lever* översätts exempelvis *DRA ALLA LEVER* från versaler, istället för *Pulling varje spak* från normalt formaterad text. Generellt är det fler engelska ord kvar vid översättning från versaler. Detta gäller även mitt översättningsmaterial från spanska; översättningen från mitt källmaterial i vanligt format innehåller inga engelska ord alls, medan samma källtext formerad till versaler genererar en översättning med tydligare spår av att delar av texten översatts via engelska. Tidsskriften *Der Spiegel* översätts exempelvis med *The Weekly* eller *The Journal*.

Redan av dessa exempel kan man sluta sig till att en ändring mellan versaler och gemener i källtexten kan ha en uppenbar påverkan på korrektheten (som blir större på vissa områden och mindre på andra). Översättningen från versaler torde dock i detta fall vara mindre begriplig: Originalet säger inget om att Indiens centralbank skulle vara delaktig i regeringen. Vidare tolkas metaforen *borrowing binge* inte som en sammansättning, vilket är längre ifrån originalet än den normalt formaterade översättningen. Därmed inte sagt att kvaliteten har blivit sämre i alla detaljer vid versalöversättningen. En möjlig slutsats är att den information som ligger i distinktionen versaler/gemener är viktig för GT.

4.4.2 Variation beroende på kontext

Översättningar av ord kan också variera beroende på om de översätta självständigt, inom en fras eller en mening (eller en ofullständig fras eller mening) eller som en del av ett helt stycke eller en hel text. På samma sätt kan GT göra felaktiga ordval vid översättningar från en mening, men ge korrekta översättningar av samma mening i en större kontext, och vice versa.

Som exempel på att ofullständiga meningar kan ge andra ordval än hela meningar kan vi använda oss av en tidigare (exempel 3, s. 9) återgiven exempelmening, men se vad som händer om vi tar bort det sista ordet, eller bara har halva ordet kvar:

- (28) a. **KT:** India's central bank is one of its best institutions.
GT: Indiens centralbank är ett av sina bästa institutioner. (28/3 2012)
- b. **KT:** India's central bank is one of its best
GT: Indiens centralbank är ett av dess bästa (28/3 2012)
- c. **KT:** India's central bank is one of its best insti
GT: Indiens centralbank är ett av dess bästa institutioner (28/3 2012)

Som vi ser är översättningen av den första exempelmeningen identisk med den översättning vi fick ut när vi tog med hela artikeln. Men tar vi bort *institutioner* från meningen väljer tjänsten plötsligt rätt pronomen för sammanhanget. Skriver vi bara in *insti* väljer tjänsten att automatiskt lägga till hela ordet *institutioner* och dessutom välja rätt pronomen. Vi får alltså en mer korrekt översättning från en källtext med ett halvt slutord!

Det kvarstår fortfarande ett genusfel i samtliga av dessa översättningar. Men också här är rätt eller fel en fråga om kontext, jfr ex. (28 a–c) med exemplen i (29):

- (29) a. **KT:** An institution
GT: En institution (26/4 2012)
- b. **KT:** A fine institution
GT: En fin institution (26/4 2012)
- c. **KT:** One institution
GT: En institution (28/3 2012)
- d. **KT:** One of its best institutions.
GT: En av sina bästa institutioner. (28/3 2012)

Tar man bort punkten i den sista av dessa exempelmeningar uppstår problemet att GT tolkar också den engelska pluralformen som en genitivform och dessutom sätter ordet i bestämd form:

- (29) e. **KT:** One of its best institutions
GT: En av sina bästa institutionernas (28/3 2012)

Nu är denna typ av genusproblematik som tidigare nämnts (4.2.1) troligen mer relaterad till korrekthet än till begriplighet. Felaktiga pronomen skulle troligen kunna vålla mer problem än i dessa givna exempel. Exempelen ovan förstärker dock också bilden av att kollokationer som står för sig själva, likt 29 a-c, blir lättare för GT att översätta korrekt.

Det finns andra exempel på att översättning av polysema ord, eller av ord med flera möjliga översättningsalternativ, tycks påverkas starkt av kontexten i GT. Skriver man in sådana ord i en fras kan man ibland se hur översättningen ändras bara efter några fler bokstäver. Nedan ett exempel på sådana förändringar vid översättning från svenska till engelska:

(30)

KT	GT 28/3 2012
Vd	CEO
Vd för	Managing Director
Vd för Ce	President of Ce
Vd för Cen	Managing Director for Central
Vd för Cent	CEO of Cent
Vd för Centr	Managing Director of Centr
Vd för Centra	CEO of Centra
Vd för Central	Managing Director of Central
Vd för Centralb	CEO of Centralb
Vd för Centralba	CEO of Centralba
Vd för Centralban	Managing Director of Central
Vd för Centralbank	Managing Director of Central
Vd för Centralbanke	CEO of Central Banke
Vd för Centralbanken	CEO of National Bank

Som vi ser vacklar översättningen av *Vd* nästan för varje ny bokstav som skrivs in. Vi ser också ett exempel där GT inte gör någon gissning om slutordet förrän hela ordet redan är på plats. Klickar man fram förslagna alternativa översättningar av *Vd* förekommer både *CEO*

och *Managing Director* från inmatningen *Vd för Centralb* och framåt. Innan dess förekommer *President* också som förslag när källtexten är *Vd fö*, *Vd för Cent* och *Vd för Central*, men inte annars.

Varken den engelska eller den spanska ekonomiartikeln i mitt källmaterial översätts annorlunda om man klistrar in hela artikeln (1749 ord) istället för det fragment på ca 200 ord jag valt att undersöka.

4.4.3 Variation vid byte av skiljetecken

I min engelska exempelartikel finns en mening som använder ett semikolon. Detta skiljetecken tycks vålla en del problem:

- (31) a. **KT:** In normal times this is a task that would furrow the brow; now that the country's boom is faltering, it risks causing a blinding headache.

GT: I normala tider är en uppgift som skulle fåra pannan, nu att landets bommen vacklar, riskerar att den kan orsaka en bländande huvudvärk. (5/2 2012)

GT gör för det första om skiljetecknet till ett komma. För det andra görs en inkorrekt tolkning av kollokationen *now that*, något som tillsammans med en otydlig syntax går ut över begripligheten.

Ersätter vi källtextens semikolon i (31) med punkt, får vi en annan och kanske något mer begriplig översättning:

- (31) b. **KT:** In normal times this is a task that would furrow the brow. Now that the country's boom is faltering, it risks causing a blinding headache.

GT: I normala tider är en uppgift som skulle fåra pannan. Nu när landets bommen vacklar, riskerar att den kan orsaka en bländande huvudvärk. (11/4 2012)

Som vi ser återstår mycket av de syntaktiska problemen, men det blir samtidigt lättare att skilja de två huvudsatserna från varandra. Vi får också en korrekt översättning av *now that*. En aspekt på skillnaden mellan semikolon och komma skulle kunna vara att engelskt semikolon ofta motsvaras av svenskt kommatecken i GT:s källmaterial.

4.4.4 Variation beroende på källspråk?

Wasmuth (2011: 11) hävdar att vissa språkpar är lättare att maskinöversätta mellan än andra. Vi kan konstatera att det mellan engelska och spanska finns en del skillnader som också leder till att problemen vad gäller korrekthet återfinns på lite olika ställen i de svenska översättningarna beroende på källspråk. Kongruensproblem med adjektiv har jag exempelvis oftare hittat i det engelska översättningsmaterialet (med reservation för att mitt material är litet). Problemen med översättningar av genitiv är ett annat exempel där spanskan utmärker sig som annorlunda eftersom spanskan uteslutande kan konstruera genitiv med preposition och inte med suffix (4.3.6).

Dessa korrekthetsrelaterade problem kan ibland påverka begripligheten (som vi sett med exempelvis spanska genitivformer). De problem som inverkar på översättningarnas begriplighet är dock likartade: polysemi, bokstavligt tolkade metaforer m.m. För båda språken tycks också mer kreativa formuleringar generellt skapa problem för tjänsten, medan konventionellt språk är lättare att översätta.

Man kunde tänka sig att språk med ett större antal homonymer eller polysema ord skulle vålla större problem vid översättning både från och till. Vidare kunde man tänka sig att två språk med mer olika grammatiska system blir mer problematiska att översätta mellan än språk som liknar varandra mer (jfr. Ping 2009, refererad i 2.1). Man kan dock inte utifrån min studie säga att översättningar från engelska håller högre kvalitet än de från spanska till svenska bara för att engelska också är ett germanskt språk.

4.4.5 Variation beroende på målspråk?

I avsnitt 3 tog jag upp Shen (2010) som hävdar att översättning med GT fungerar speciellt bra mellan vissa språk och att engelskan ofta tenderar att vara exceptionellt bra som målspråk. Shen menar dock att språkparet engelska och spanska är ett undantag från denna tendens, där översättningarna tycks fungera bättre från engelska till spanska. Till detta anger han som möjlig orsak att många latinamerikanska länder erbjuder engelska översättningar av centrala offentliga dokument.

Med spanska som källspråk kan vi dock testa hur bra svenska är som målspråk jämfört med engelska. Om översättning till engelska fungerar sämre från spanska än från jämförbara stora EU-språk tycks det dock i vissa fall fungera bättre som målspråk än svenska. Jämför man ett wikipedia-uppslag på spanska, översatt till svenska respektive engelska (se bilaga 2)

tycks den svenska översättningen få något fler problem av det slag som tagits upp i både 4.2 (korrekthet) och 4.3 (begriplighet):

- (32) 1 **KT:** En 1813 le fue concedido el título honorífico de Libertador por el Cabildo de Mérida en
3 Venezuela que, tras serle ratificado en Caracas ese mismo año, quedó asociado a su nombre.
5 Los problemas para llevar adelante sus planes fueron tan frecuentes que llegó a afirmar de sí
7 mismo que era "el hombre de las dificultades" en una carta dirigida al general Francisco de
9 Paula Santander en 1825.
- 7 Participó en la fundación de la Gran Colombia, nación que intentó consolidar como una gran
9 confederación política y militar en América, de la cual fue Presidente. Bolívar es considerado
11 por sus acciones e ideas el "Hombre de América" y una destacada figura de la Historia
13 Universal, ya que dejó un legado político en diversos países latinoamericanos, algunos de los
cuales le han convertido en objeto de veneración nacionalista. Ha recibido honores en varias
partes del mundo a través de estatuas o monumentos, parques, plazas, etc. Así mismo, sus ideas
y posturas política-sociales dieron origen a una corriente o postura llamada Bolivarianismo.
- 1 **GT, sv:** År 1813 tilldelades han hederstiteln Liberator av kommunfullmäktige i Merida i
3 Venezuela, Caracas när de har ratificerats under det året blev förknippad med hans namn.
5 Problemen för att genomföra sina planer var så täta att till och med sagt själv att det var
7 "mannen i svårigheter" i ett brev adresserat till General Francisco de Paula Santander i 1825.
9 Deltog i grundandet av Gran Colombia, en nation som försökt att befästa som en politisk och
11 militär konfederation i Amerika, där han var VD. Bolivar anses av sina handlingar och tänka
"Man of America" och en ledande figur i världshistorien som den lämnade ett politiskt arv
i flera latinamerikanska länder, av vilka några har blivit föremål för vördnad nationalist. Han
har fått utmärkelser från olika delar av världen genom statyer och
monument, parker, torg etc.. Dessutom gav hans idéer och politiska och sociala
positioner upphov till en ström är eller Bolivarianism. (4/5 2012)
- 1 **GT, eng:** In 1813 he was awarded the honorary title of Liberator by the City Council of Merida
3 in Venezuela, Caracas after being ratified in that year became associated with his name. The
5 problems to carry out his plans were so frequent that even said himself that it was "the man of
7 the difficulties" in a letter addressed to General Francisco de Paula Santander in 1825.
9 Participated in the foundation of Gran Colombia, a nation that tried to consolidate as apolitical
11 and military confederation in America, of which he was President. Bolivar is considered by
his actions and thinking "Man of America" and a leading figure in world history as it left
a political legacy in several Latin American countries, some of which have become the object
of veneration nationalist. He has received honors from various parts of the world
through statues and monuments, parks, plazas, etc.. Moreover, his ideas and political and
social positions gave rise to a current pose or Bolivarianism. (4/5 2012)

I det första stycket i (32) har den svenska översättningen ett kvarvarande engelskt ord (*Liberator*, r.1), lexikala polysemiproblem (*kommunfullmäktige*, r.1, *täta* problem, r. 3), pronominella problem (*de*, r.2, *det*, r. 3) och förlorad information ("att [han] till och med sagt [om] sig själv...", r. 3).

Den engelska översättningen har samma problem med förlorad information och pronominella fel, men tycks hantera de lexikala polysemiproblemen bättre (t.ex. genom det mer icke-nationsbundna *City Council*, r.1, och *frequent*, r.3). Den engelska översättningen översätter också en spansk gerundiumform effektivt med en engelsk ing-form (*after being ratified*, r.2).

I stycke 2 har både svenskan och engelskan samma syftningsfel i den första raden. I den andra raden har dock engelskan en fördel av att översätta *presidente* med *president* istället för den svenska översättningens *VD*. Följande menings konstruktion, *Bolivar anses av sina handlingar...* tycks mindre begriplig än engelskans *is considered by his actions*.

Det är dock inte entydigt i detta exempel att engelskan har en fördel mot svenskan: I exemplet avslutningsmening talas om en *corriente* (alltså en politisk strömning), vars engelska översättning *current* vållar problem. Ordet kan nämligen tolkas som homonymen med betydelsen 'nuvarande'. I detta sammanhang torde den svenska översättningen *ström* vara något mer funktionell.

Shen (2010) ger visserligen inte någon statistik som inkluderar svenskan, men att döma av hans jämförelser mellan språkpar där engelskan ingår, tillsammans med mina iakttagelser i ex. (32), kan man få uppfattningen att engelskan generellt skulle kunna hålla högre kvalitet som målspråk än svenskan. Då engelskan både är ett större språk, och således har ett större översatt material att tillgå, och har funnits som givet käll- och målspråk för GT sedan programmets första dagar, medan svenskan tillkommit senare, kan en sådan slutsats tyckas logisk.

4.4.6 Sammanfattning av de viktigaste resultaten

För det första kan man utifrån min studie sluta sig till att GT har både förtjänster och begränsningar och då både när det gäller korrekthet och begriplighet. De observationer jag gjort kan ofta på ett eller annat sätt relateras till GT:s statistiska tillvägagångssätt: Konventionella standardfraser har större sannolikhet att förekomma i en korpus än mer ovanliga ordkombinationer. Ibland (men inte alltid) kan förekomsten av vissa grammatiska konstruktioner och ordval också understödjas av att det valda alternativet är mer vanligt förekommande i Googles sökmotor. Centrala problem för googleöversättningars begriplighet tycks kretsa kring polysemi, metaforer och i vissa fall grammatisk struktur (exempelvis prepositioner, när de används till olika saker på olika språk).

GT:s lösningar tycks också variera mycket beroende på form och kontext. Omformar man källtexten till endast versaler, eller tar bort eller ändrar ett skiljetecken, kan innehållet i översättningen påverkas. I långa, komplexa meningar kan det (i vissa fall) göra texten mer korrekt/begriplig att ta bort delar av meningen.

En stark sida hos GT tycks vara att översätta konventionella standardfraser, medan t.ex. ovanliga, kreativa metaforer tycks innebära problem för tjänsten. Jag finner inte stöd för att något av de båda källspråken skulle ge högre kvalitet vid GT-översättning till svenska.

5. Slutdiskussion

Min undersökning har visat att GT:s resultat framför allt varierar beroende på form och kontext. De utdrag av översättningar jag har detaljstuderat ger dock inte tydliga svar om hur GT:s kvalitet kan variera beroende på textgenre (eftersom de tillhör samma genre). En del kan dock sägas utifrån dessa artiklar, som relaterar till genre. Enligt Wasmuth finns förutom språkpar som är mer eller mindre lätta att maskinöversätta mellan, också ”olika texttyper som är mer eller mindre svåröversatta. Kommersiella brev och tekniska eller ekonomiska texter är inga problem för ett maskinöversättningssystem medan poesi eller marknadsföringstexter inte fungerar alls.” (Wasmuth 2011: 11).

Utan att kunna testa detta påstående fullt ut kan vi konstatera att de genrer som av Wasmuth beskrivs som lätthanterliga vid maskinell översättning kännetecknas av fackspråk och stringens snarare än kreativitet. Detta innebär i många fall en begränsad användning av levande metaforer, och i synnerhet av ordlekar (har vi att göra med tekniska instruktioner förväntas de av naturliga skäl inte ha dubbel betydelse). I poesi och marknadsföringstexter är det samtidigt ofta vanligt och inte sällan centralt med just ordlekar och dubbla betydelser. Således kan genrekonventioner relateras till förväntad förekomst av exempelvis framträdande polysemi och levande metaforer.

Som en utveckling av detta resonemang kan vi också se att den engelska ekonomiartikel jag använt som källmaterial för min analys inte befinner sig i någon av dessa två ytterligheter, utan tillhör en genre som blandar nationalekonomiska termer med utpräglad kreativa, reportageliknande skildringar. Den spanska artikeln använder däremot inte denna typ av kreativa metaforer och skildringar av omgivningen. Här handlar det mer om att objektivt återge ett händelseförlopp, även om ett smärre antal metaforer (t.ex. *skakar om*) används.

En slutsats av detta är att en text inte behöver vara lätthanterlig för GT bara för att den handlar om exempelvis ekonomi. Istället tycks skiljelinjen gå mellan mer funktionella och

mer kreativa texter. Ju mer kreativ text, desto svårare att bedöma dess betydelse utifrån statistik. Detta kan också kopplas samman med tidigare iakttagelser i denna uppsats om GT:s större kapacitet att översätta mer konventionella kollokationer, ”standardfraser”, såväl när det handlar om metaforer som polysemi eller kongruens.

Av mina resultat kan man bland annat utröna att det går att få ut mer av tjänsten än man först ser, om man vet hur. Förhoppningsvis kan denna kunskap användas för att nyttja tjänsten till fullo. Mina resultat torde också kunna ge en bild av vilka drag i källtexter som kan tänkas vålla mer och mindre problem vid översättning, vilket också kan vara värdefull information för tjänstens användare. Har man en god förståelse av engelska visar mina iakttagelser också att man kan ha fördel av att välja detta målspråk framför svenska, då dess översättningar tycks mer korrekta och begripliga. Med enkla googlesökningar och översättningar av standardfraser hoppas jag ha kunnat ge en bild av tjänstens statistiska approach.

Det finns många frågor kvar efter denna studie. Fler språkpar skulle behöva studeras för att på allvar testa teserna om käll- och målspråkens påverkan. Framför allt skulle fler utomeuropeiska språk behöva ingå i jämförelserna. Min studie begränsar sig också till att ge en prognos om vilka fenomen som kan tänkas vålla problem. Ett förslag till framtida forskning är därför en empirisk studie av människors förståelse av googleöversättningars innehåll. Ett annat förslag till en vidare studie kunde vara att inkludera fler textgenrer för att mer ingående pröva tesen om varierade resultat beroende på genre.

Slutligen har min studie en begränsning som tycks gemensam med de flesta studier om statistisk maskinöversättning: Eftersom GT:s korpus och teknik tycks uppdateras kontinuerligt är den till viss utsträckning en färskvara. Jag anser den dock vara en funktionell studie att såväl utvinna information från som bygga vidare på för framtida forskning.

6. Litteraturförteckning

Primärlitteratur

Osigh. 2012: The Reserve Bank of India – Pulling every lever. I: *The Economist* 3/2 2012.
London: The Economist Group (red. Micklethwait, John).
<http://www.economist.com/node/21546004> (3/2 2012)

Osigh. 2012: Berlín estudia subvencionar a Grecia a través del consumo eléctrico alemán. I: *El Economista* 5/2 2012. Spanien (ospec. ort): Editorial Ecoprensa (red. Ayora, Amador G.).

Björn Moberg, våren 2012

<http://www.eleconomista.es/economia/noticias/3723522/02/12/-Berlin-estudia-subvencionar-a-Grecia-a-traves-del-consumo-electrico-aleman.html> (5/2 2012)

Wikipedia (Uppslagsord: Simón Bolívar. Språk: spanska). Wikimedia foundation.

http://es.wikipedia.org/wiki/Sim%C3%B3n_Bol%C3%ADvar (15/5 2012)

Sekundärlitteratur

Adams, Tim, 2010: Can Google break the computer language barrier? I: *The Guardian* 19/12 2010. London: Guardian Media Group (red. Rusbridger, Alan).

<http://www.guardian.co.uk/technology/2010/dec/19/google-translate-computers-languages> (7/5 2012)

Domeij, Rickard, 2011: Datorn som god översättare i snäv mening. I *Språktidningen* 3: 44-49. Stockholm: Språktidningen i Sverige AB (red. Hadenius, Patrick).

Hartley, Tony, 2009: Technology and translation. I: *The Routledge Companion to Translation studies* 106-127. New York: Routledge (red. Munday, Jeremy).

Helft, Miguel, 2010: Google's computing power refines translation tool. I *The New York Times* 8/4 2011. New York: New York Times Company (red. Sulzberger Jr, Arthur O).

<http://www.nytimes.com/2010/03/09/technology/09translate.html> (7/5 2012)

Kingsley, Jeremy, 2011: It already speaks 57 languages as well as a 10-year-old. How good can it get? I *Slate* 31/10 2011, Washington: Washington Post Company (red. Plotz, David).

http://www.slate.com/articles/technology/technology/2011/10/google_translate_will_google_s_computers_understand_languages_be.single.html?utm_medium=referral&utm_source=pulse enews (7/5 2012)

López, Adam, 2008: Statistical Machine Translation. I: *ACM Computing Surveys* 40 (3): 1-49. New York: ACM.

<http://www.cs.jhu.edu/~alopez/papers/survey.pdf> (5/12 2011).

Björn Moberg, våren 2012

Lundquist, Lita, 2005: *Oversættelse. Problemer og strategier set i tekstlingvistisk og pragmatisk perspektiv*. Frederiksberg: Samfundslitteratur

Ping, Ke, 2009: Machine Translation. I *Routledge Encyclopedia of Translation Studies*. New York: Routledge (red. Baker, Monica, Saldanha, Gabriela, 2:a uppl.).

Shen, Ethan, 2010: Comparison of online machine translation tools. I *Tcworld* juni 2010. Tyskland (ospec. Ort), Tekom.

<http://web.archive.org/web/20110210034229/http://tcworld.info/index.php?id=175> (7/5 2012)

Wasmuth, John, 2011: *CAT-verktyg: Hot eller möjlighet?* Lund: Lunds Universitet.

<https://lup.lub.lu.se/luur/download?func=downloadFile&recordId=1975380&fileId=1975400>

Bilagor

Bilaga 1

The Economist (2012), omformad till versaler. GT 26/3:

Reserve Bank of India

DRA ALLA LEVER

INDIEN centralbank är en av de bästa institutionerna. DET medverkar även i en regering som lånar SUPKALAS

Februari 4TH 2012 | Mumbai | från den tryckta upplagan

EN av de förmåner med att vara chef för Reserve Bank of India (RBI) är användningen av kolonial bungalow på Carmichael Road, POSH STREET som väver längs en

ås i SOUTH Mumbai. PÅ EN SIDA LIVE några av

Indiens rikaste industriledare, ModernaPharaohs MED Flashy ARKITEKTUREN smaker. Å andra sidan Ett stenkast DOWN en klippa, är en liten SLUM - ett monument desperation och regeringars misslyckanden. Båda GRANNAR ÄR EN DEL AV 1,2 miljarder invånare att Indiens centralbanken måste se upp med. I normala tider detta är en uppgift som SKULLE fåra panna; NU att landets BOOM vacklar, riskerar att den kan orsaka ett bländande Huvudvärk.

Att döma av siffrorna ÄR RBI Bland världens bästa centralbankerna. ITS RECORD på att balansera tillväxt och inflation är anständig nog (se diagram 1). Sedan 1995 har grossistpriserna har stigit med i genomsnitt 6% per år, inte för långt FRÅN RBI S Comfort Zone på cirka 5%. Tillväxten har genomsnitt 7% per år.

El Economista (2012), omformad till versaler, GT 26/3:

BERLIN GREKLAND STUDIE subventionera genom den tyska STRÖMFÖRBRUKNING
2012/05/02 - 12:55

0

REGERING förbundskanslern, Angela Merkel, studerar stödberättigande Greklands krisdrabbade ekonomin, fast i skuldkrisen, GENOM användning av el i Tyskland som varje vecka Der Spiegel.

Att uppmuntra investeringar i VÄXTER sol- eller vindkraftverk, tyska regeringen subventioner för förnyelsebar energi kan också dra nytta levererar till grekiska järnvägsnätet europeiska mervärde Weekly.

BERLIN STUDIE BEVILJAT public STÖD kreditinstitut för återuppbyggnad (KfW) investeringar i Grekland för förnybar energi.

The Weekly planer ITS INFORMATION KOMMISSION Europeiska unionens vägnar ett nytt program för att främja ekonomisk tillväxt i Grekland.

Tillagt att dock både ökningen mellan Kritiskt för Aten för Din attityd tveksam När man arbetar med finansiella kris som skakar landet.

The Journal Hittar trojkan EU-kommissionen, Europeiska centralbanken BANK OCH Internationella valutafonden vara nödvändig för att ATEN MINSKA arbetskraftskostnaderna mellan 15 och 20 procent.

Minimilön i Grekland fortsätta till 50 procent högre nivå Portugal, som härrör från den förväntade denna rapport Denna vecka GRUPP inspektörer från den trojkan.

Bilaga 2

Ur *Wikipedia* (2012), "Simón Bolívar" 15/5 2012

Simón José Antonio de la Santísima Trinidad Bolívar y Palacios Ponte y Blanco, mejor conocido como Simón Bolívar, (Caracas, 24 de julio³ de 1783 — Santa Marta, República de Colombia, 17 de diciembre de 1830) fue un militar y político venezolano de la época pre-republicana de la Capitanía General de Venezuela; fundador de la Gran Colombia y una de las figuras más destacadas de la emancipación americana frente al Imperio español. Contribuyó de manera decisiva a la independencia de las actuales Bolivia, Colombia, Ecuador, Panamá, Perú y Venezuela.

En 1813 le fue concedido el título honorífico de Libertador por el Cabildo de Mérida en Venezuela que, tras serle ratificado en Caracas ese mismo año, quedó asociado a su nombre. Los problemas para llevar adelante sus planes fueron tan frecuentes que llegó a afirmar de sí mismo que era "el hombre de las dificultades" en una carta dirigida al general Francisco de Paula Santander en 1825.

Participó en la fundación de la Gran Colombia, nación que intentó consolidar como una gran confederación política y militar en América, de la cual fue Presidente. Bolívar es considerado por sus acciones e ideas el "Hombre de América" y una destacada figura de la Historia Universal, ya que dejó un legado político en diversos países latinoamericanos, algunos de los cuales le han convertido en objeto de veneración nacionalista. Ha recibido honores en varias partes del mundo a través de estatuas o monumentos, parques, plazas, etc. Así mismo, sus ideas y posturas política-sociales dieron origen a una corriente o postura llamada Bolívarianismo.

GT (15/5 2012) , **sv**:

Simon Jose Antonio de Bolivar och Trinity Ponte Palacios y Blanco, mer känd som Simon Bolivar (Caracas, 24 julio³, 1783 - Santa Marta, Colombia, december 17, 1830) var en venezuelansk militär och den politiska pre-republikanska av kaptengeneralsekreterare Venezuela, grundare av Gran Colombia och en av de ledande figurerna i American Independence mot det spanska imperiet. Avgörande sätt

bidragit till oberoende av föreliggande Bolivia, Colombia, Ecuador, Panama, Peru och Venezuela.

År 1813 tilldelades han hederstiteln Liberator av kommunfullmäktige i Merida i Venezuela, Caracas när de har ratificerats under det året blev förknippad med hans namn. Problemen för att genomföra sina planer var så täta att till och med sagt själv att det var "mannen i svårigheter" i ett brev adresserat till General Francisco de Paula Santander i 1825.

Deltog i grundandet av Gran Colombia, en nation som försökt att befästa som en politisk och militär konfederation i Amerika, där han var VD. Bolivar anses av sina handlingar och tänka "Man of America" och en ledande figur i världshistorien som den lämnade ett politiskt arv i flera latinamerikanska länder, av vilka några har blivit föremål för vördnad nationalist. Han har fått utmärkelser från olika delar av världen genom statyer och monument, parker, torg etc.. Dessutom gav hans idéer och politiska och sociala positioner upphov till en ström är eller Bolivarianism.

GT (15/5 2012), eng:

Simon Jose Antonio de Bolivar and Trinity Ponte Palacios y Blanco, better known as Simon Bolivar (Caracas, julio3 24, 1783 - Santa Marta, Republic of Colombia, December 17, 1830) was a Venezuelan military and the political pre-republican of the Captaincy General of Venezuela, founder of Gran Colombia and one of the leading figures of American Independence against the Spanish Empire. Contributed decisively to the independence of the present Bolivia, Colombia, Ecuador, Panama, Peru and Venezuela.

In 1813 he was awarded the honorary title of Liberator by the City Council of Merida in Venezuela, Caracas after being ratified in that year became associated with his name. The problems to carry out his plans were so frequent that even said himself that it was "the man of the difficulties" in a letter addressed to General Francisco de Paula Santander in 1825.

Participated in the foundation of Gran Colombia, a nation that tried to consolidate as a political and military confederation in America, of which he was President. Bolivar is considered by his actions and thinking "Man of America" and a leading figure in world history as it left a political legacy in several Latin American countries, some of which

have become the object of veneration nationalist. He has received honors from various parts of the world through statues and monuments, parks, plazas, etc.. Moreover, his ideas and political and social positions gave rise to a current pose or Bolivarianism.

Bilaga 3

Utdrag ur *The Economist* (2012) i spalt 1, googleöversättning av texten i spalt 2, förf. anmärkningar i spalt 3.

The economist (2012): “The Reserve Bank of India – Pulling every lever”

Pulling every lever

India’s central bank is one of its best institutions. It is also complicit in a government-borrowing binge

Feb 4th 2012 | *MUMBAI* | from the print edition

ONE of the perks of[1] being governor of[2] the Reserve Bank of India (RBI) is the use of a colonial bungalow on Carmichael Road, a posh street that weaves along a ridge in south Mumbai. On one side live some of India’s richest industrialists, modern-day pharaohs with flashy architectural tastes. On the other, a stone’s throw down[3] a cliff, is a small slum—a monument to[4] desperation and government failure.

GT5/2 2012: Reserve Bank of India

Pulling[5] varje spak

Indiens centralbank är ett[6] av sina[7] bästa institutioner. Det är delaktig[8] även[9] i ett statligt lån hetsätning[10]

4 feb 2012 | Mumbai | från den tryckta upplagan

[11]EN av de förmåner[12] med att vara guvernör[13] Reserve Bank of India (RBI) är användningen av[14] en kolonial bungalow på Carmichael Road, en posh[15] gata som väver[16] längs en ås i södra Mumbai[17]. På ena sidan bor några av Indiens rikaste industrimän, moderna faraoner med flashiga arkitektoniska smak[18]. Å andra sidan[19] ett stenkast ner en klippa, är en liten slum-ett monument desperation och regeringen misslyckande[20].

1: Korrekt översatt preposition. 2: Utelämnad engelsk preposition. 3: Tolkas som 'ner' och inte 'nedför', vilket gör meningen ogrammatisk. 4: Utelämnad engelsk preposition. 5: engelskt ord kvar. 6: Fel genus. 7: Fel possessivt pronomen. 8: Kongruensfel. 9: Felplacerat Satsadverbial. 10: Misstolkad engelsk sammansättning. Metaforen tolkas bokstavligt. 11: Korrekt konstruktion. 12: Misstolkad engelsk artikel. 13: Korrekt översatt ing-form. *Governor* översätts 'guvernör'. 14: nominalisering som normalt skulle lösas upp vid översättning. 15: Engelskt ord kvar. 16: Metaforen tolkas bokstavligt. 17: Korrekt syntaktisk konstruktion bortsett från de fel som finns i meningens fundament. 18: Kongruensfel, i övrigt är meningen korrekt. 19: Fel preposition. 'På' tolkas här som 'å'. 20: Misstolkad engelsk sammansättning.

Forts. kommentarer på utdrag ur *The Economist* (2012)

Both sets of neighbours are part of[21] the 1.2 billion population that India's central bank must look out for. In normal times this[22] is a task that would furrow the brow; now that the country's boom is faltering, it risks causing a blinding headache.

Judging by the numbers, the RBI is among the world's best central banks. Its record on[23] balancing growth and inflation is decent enough (see chart 1). Since 1995 wholesale prices have risen by[24] an average of 6% a year, not too far from the RBI's comfort zone of[25] about 5%. Growth has averaged 7% a year

Båda grannar är en del av de 1,2 miljarder invånare att[26] Indiens centralbank måste se upp för[27]. I normala tider är en uppgift som skulle fåra pannan,[28] nu att[29] landets bommen[30] vacklar, riskerar att den kan[31] orsaka en bländande huvudvärk.

Att döma av siffrorna är RBI bland världens bästa centralbanker[32]. Dess[33] resultat när det gäller balansering tillväxt [34]och inflation är anständiga[35] nog (se diagram 1). Sedan 1995 har grossistpriserna har[36] stigit med i genomsnitt 6% per år, inte alltför långt från Indiens centralbank[37] komfort zon ca 5%. Tillväxten har i genomsnitt [38]7% per år.

21: Denna preposition översätts. 22: Översatt demonstrativt pronomen. 23: Korrekt översatt preposition. 24: Korrekt översatt preposition. 25: Utelämnad preposition. 26: Engelsk relativ bisats tolkad som att-sats. 27: *Look out for* tolkas bokstavligt. 28: Semikolon blir komma; den efterföljande syntaktiska engelska meningen tolkas som delar av den förra syntaktiska meningen. 29: Temporär bisats tolkad som att-sats. 30: Huvudordet i en genitivfras får bestämd form. Metaforen (*bommen*) tolkas bokstavligt. 31: Extrainsatt verb. Den engelska ing-formen tolkas som en att-sats (med 'att' som subjunktion och inte infinitivmärke). 32: Korrekt mening. Här klarar tjänsten av både ing-formen och genitivkonstruktionen. 33: Korrekt possessivt pronomen. 34: Utelämnad svensk preposition som hade behövts för att göra en grammatisk översättning av ing-formen. I övrigt bra översättning med *när det gäller*. 35: Korrekt kongruensbörjning. 36: Verbet dubbleras. 37: Utelämnat genitiv. Här översätts RBI till svenska, till skillnad från i resten av texten. 38: Utelämnat huvudverb.

Utdrag ur *El Economista* (2012) i spalt 1, googleöversättning av texten i spalt 2, förf. anmärkningar i spalt 3.

El Economista (2012): “Berlín estudia subvencionar a Grecia a través del consumo eléctrico alemán”

Berlín estudia subvencionar a Grecia a través del consumo eléctrico alemán

5/02/2012 - 12:55

El Gobierno de la canciller federal, Angela Merkel, estudia la posibilidad de subvencionar la debilitada economía de Grecia, sumida en la crisis de la deuda, a través del consumo de electricidad en Alemania, según el semanario *Der Spiegel*.

Para incentivar las inversiones en plantas solares o parques eólicos, las **subvenciones públicas alemanas a las energías renovables** podrían beneficiar también a empresas griegas que suministren electricidad[1] a la red europea, añade el semanario.

Berlín estudia también la **concesión de ayudas** por parte del público Instituto Crediticio para la Reconstrucción (KfW) a las inversiones en Grecia para las energías renovables.

GT 5/2 2012:

Berlin studerar att subventionera Grekland genom den tyska elförbrukningen[2]

2012/05/02 - 12:55

Regeringen i förbundskanslern[3] är[4] Angela Merkel överväger[5] att subventionera[6] den krisdrabbade ekonomin i Grekland[7], fast i skuld kris[8], genom konsumtion av el i Tyskland, enligt veckans[9] *Der Spiegel*.

För att uppmuntra investeringar i solenergi växter[10] eller vindkraftverk, de tyska offentliga subventioner[11] till förnybar energi skulle[12] också kunna gynna grekiska företag som levererar till det europeiska nätverket, tillade vecka[13].

Berlin är också överväger[14] att bevilja[15] offentligt stöd av Credit Institutet[16] för återuppbyggnad (KfW) för investeringar i Grekland för förnybar energi.

1: vad som levereras utelämnas i översättningen. 2: Korrekt mening bortsett från felaktigt ordval. Alla prepositioner korrekt översatta. 3: Misstolkad spansk genitivpreposition. 4: Felaktigt insatt kopula 5: här översätts samma ord (*estudia*) korrekt. 6: Korrekt insatt infinitivmärke. 7: Här är en likadan översättning som i [2]. Denna gång är den dock inte felaktig. 8: Participfras som normalt skulle lösas upp vid översättning. 9: *El semanario* tolkas som *veckans*. 10: *Planta* får betydelsen 'växt'. polysemi 11: Missad dubbel bestämning. 12: Missad V2-regel. I övrigt korrekt mening. De bestämda artiklar som hoppas över hade kunnat utelämnas av en mänsklig översättare. 13: *Semanario* tolkas som *vecka*. 14: Insatt kopula. Om kopulaverbet tas bort bryter meningen mot V2-regeln. I övrigt näst intill korrekt mening. De utelämnade bestämda artiklarna stör inte. 15: Tjänsten klarar av att lösa upp en nominalisering på korrekt sätt. 16: Tjänsten går via engelskan och tolkar den engelska sammansättningen som två ord. Ett av orden förblir i sin engelska form.

Forts. kommentarer på utdrag ur *El Economista* (2012)

El semanario basa sus informaciones en planes de la Comisión de la Unión Europea para un nuevo programa de fomento del crecimiento económico en Grecia. Agrega que, sin embargo, aumentan entre tanto las críticas hacia Atenas por su dubitativa actitud a la hora de afrontar la crisis financiera que[17] sacude al país. La revista revela que la troika formada[18] por la Comisión de la UE, el Banco Central Europeo y el Fondo Monetario Internacional consideran imprescindible que Atenas reduzca los costes laborales entre un 15 y un 20 por ciento. El salario mínimo en Grecia sigue estando un 50 por ciento por encima del nivel de Portugal, según se desprende del informe que previsiblemente presentará esta semana el grupo de inspectores de la troika.
Den veckovisa[19] bygger[20] sin information om planer på[21] EU-kommissionen för ett nytt program för att främja[22] den ekonomiska tillväxten i Grekland. Han[23] tillägger dock att öka[24] kritik[25] från både Aten för sin tveksamma inställning i hanteringen av finanskrisen sveper över landet[26]. Tidningen avslöjar att trojkan av EU-kommissionen, Europeiska centralbanken och IMF anser att det är viktigt[27] att Aten minska[28] arbetskostnaderna med[29] 15 till 20 procent. Minimilönen i Grekland är fortfarande 50 procent över nivån i Portugal, enligt den rapport som väntas presentera denna vecka en grupp av inspektörer från trojkan[30].
17: Översatt subjunktion. 18: Översatt particip. 19: <i>El semanario</i> tolkas som <i>den veckovisa</i>. 20: Död spansk metafor tolkas bokstavligt. 21: Missstolkad preposition. 22: Korrekt upplösning av en nominalfras. 23: Referenten <i>Den veckovisa</i> ges den identiska anaforen <i>han</i>. 24: Spansk bisats tolkas som infinitivfras. 25: Korrekt singularöversättning av spansk plural, <i>las críticas</i>. 26: Tjänsten väljer en annan metafor än KT, men den fungerar lika väl. 27: Korrekt skapad bisats transformerad från ett spanskt adjektiv, <i>imprescindible</i>. 28: Spansk konjunktivform tolkas som antingen imperativ eller infinitiv. 29: Korrekt insatt preposition. 30: Subjektet placeras sist, V2-brott. Annars hade meningen varit korrekt. Sammanfattningsvis är många av de spanska prepositionerna korrekt översatta och tjänsten klarar av att göra flera korrekta transformationer.