

**Computerondersteund vertalen tussen twee
kleinere talen en één grotere en één kleinere taal**

*Diplomová práce navazujícího magisterského studia
nizozemské a anglické filologie*

Lucie Kadeřábková

Vedoucí práce: **doc. dr. Wilken Engelbrecht, cand. litt.**

OLOMOUC 2012

Prohlašuji, že jsem svou diplomovou práci vypracovala samostatně a uvedla v ní veškerou literaturu, kterou jsem použila.

Ik verklaar dat ik mijn scriptie alleen geschreven heb en alle vakliteratuur en bronnen die ik gebruikt heb in de literatuurlijst heb vermeld.

V Olomouci, dne _____ 2012

Lucie Kadeřábková

Dankbetuiging

Ik wil graag van harte mijn dank betuigen aan doc. dr. Wilken Engelbrecht, cand. litt. voor zijn begeleiding en zijn waardevolle adviezen.

Lijst met afkortingen

bijv.	bijvoorbeeld
CAT	Computer Aided (Assisted) Translation
CS	Tsjechisch
d.w.z.	Dat wil zeggen
EN	Engels
enz.	enzovoorts
JTP	Jednota tlumočnicků a překladatelů
LISA	<i>Localisation</i> Industry Standards Association
MT	Machine Translation
NL	Nederlands
OCR	Optical Character Recognition
SK	Slowaaks
TAUS	Translation Automation User Society
TGB	Translator group buying (op ProZ.com)

Lijst van afbeeldingen, figuren en tabellen

Afbeelding 1 *Syntactische boom I*

Afbeelding 2 *Syntactische boom II*

Afbeelding 3 *Het proces van vertalen met een CAT tool*

Afbeelding 4 *Exact match*

Afbeelding 5 *Fuzzy match*

Afbeelding 6 *WinAlign*

Afbeelding 7 *MS Excel terminologielijst*

Afbeelding 8 *MS Word terminologielijst*

Afbeelding 9 *Google Toolkit – het proces van vertalen*

Afbeelding 10 *Globalisering, lokalisatie en vertalen*

Afbeelding 11 *Een voorbeeld machine vertaling van Babel Fish*

Afbeelding 12 *Een voorbeeld machine vertaling van Google Translate*

Figuur 1 *Vertaalgeheugen onderzoek*

Figuur 2 *HaveyoutriedGoogle'sTranslator Toolkit?(ProZ)*

Figuur 3 *What do you think about “crowdsourced“ translation?(ProZ)*

Figuur 4 *Aantal vertalers*

Figuur 5 *Talencombinaties (ProZ)*

Figuur 6 *Programma's*

Figuur 7 *Programma's (JTP)*

Figuur 8 *Aantal programma's*

Figuur 9 *Aantal programma's (ProZ)*

Figuur 10 *Reden(en) om een CAT-tool te kiezen*

Figuur 11 *Is professioneel vertalen nog mogelijk zonder CAT tools?*

Figuur 12 *Do you think CAT tool knowledge is essential for translator?(ProZ)*

Figuur 13 *Do you use a CAT tool (ProZ)*

Figuur 14 *Do you think CAT tools improve the quality of translations? (ProZ)*

Figuur 15 *Aantal vertalers die met CAT tools werken*

Figuur 16 *Aantal vertalers die met CAT tools werken II*

Figuur 17 *Voordelen van de programma's*

- Figuur 18** *Glossarium vs. Vertaalgeheugen*
- Figuur 19** *Glossarium*
- Figuur 20** *Vertaalgeheugen*
- Figuur 21** *Opleiding*
- Figuur 22** *Opleiding van de vertalers*
- Figuur 23** *Do you have bachelor's or higher degree in translation? (ProZ)*
- Figuur 24** *Do universities prepare translators / interpreters sufficiently for the life in the industry?*
- Figuur 25** *Are translators with a translation degree more competent? (ProZ)*
- Figuur 26** *Soorten teksten*
- Figuur 27** *Specialisatie*
- Figuur 28** *Hoe gespecialiseerd*
- Figuur 29** *Aantal vakgebieden*
- Figuur 30** *Zijn CAT tools nuttig?*
- Figuur 31** *Zijn CAT tools nuttig?II*
- Figuur 32** *Lokalisatie*
- Figuur 33** *Lokalisatie II*
- Figuur 34** *Are you working in the localisation industry?(ProZ)*
- Figuur 35** *Machine Translation: a translator's friend or foe?(ProZ)*
- Figuur 36** *Machine Translation: opinie van de vertalers*
- Figuur 37** *Do you use machine translation to do a first draft? (ProZ)*
- Figuur 38** *Will machine translation ever replace human translators? (ProZ)*
-
- Tabel 1** *Idiomen*
- Tabel 2** *Gelijke metaforen*
- Tabel 3** *Afwijkende metaforen*
- Tabel 4** *Aangeboden producten SDL Trados*
- Tabel 5** *Aangeboden producten Wordfast*
- Tabel 6** *Aangeboden producten Déjà Vu*
- Tabel 7** *Aangeboden producten MemoQ*
- Tabel 8** *Prijs van de programma's*
- Tabel 9** *Systeembodigheden*
- Tabel 10** *Testperiode*

Tabel 11 *Ondersteunde bestandsformaten*

Tabel 12 *Verschillen in datumformaat*

Tabel 13 *Verschillende lengte van de vertaling en de originele tekst*

Tabel 14 *Verkochten licenties op ProZ*

Tabel 15 *Programma preferenties*

Tabel 16 *Is volgens u vertalen voor de zakelijke markt nogmogelijk zonder CAT-tools?*

Tabel 17 *Vier belangrijkste voordelen*

Tabel 18 *Vier belangrijkste nadelen*

Tabel 19 *Wat vindt u belangrijker / Wat gebruikt u vaker – glossarium of vertaalgeheugen?*

Tabel 20 *Opleiding van de vertalers*

Tabel 21 *Soorten teksten*

VOORWOORD

Het hoofdthema van deze scriptie is computerondersteund vertalen. Ik wil hier graag mijn motivatie uitleggen om dit onderwerp te kiezen. Hoewel ik studeer taal- en letterkunde, is vertalen een belangrijk onderdeel van mijn studie. Ik doe het graag omdat het een praktisch en creatief onderdeel van mijn studie is en omdat het één van de mogelijke toekomstige banen voor een linguïst is. Om je broodwinning te verkrijgen met literair vertalen is vooral voor een jonge vertaler moeilijk, daarom gaat men meestal met commerciële vertalingen beginnen en wel voornamelijk handleidingen, economische of juridische teksten. Bijna alle vertaalbureaus en veel klanten verwachten van hun vertalers kennis van zogenaamde CAT tools – programma's voor computer ondersteund vertalen. Kennis van vertaalprogramma's lijkt noodzakelijk om een carrière als een vertaler te beginnen. Zulke vaardigheden ontbreken vaak echter aan onze universiteiten, zelfs bij studierichtingen die op vertalen gericht zijn.

In 2009 heb ik via het Erasmusprogramma aan de Hogeschool Gent gestudeerd en ik heb daar de mogelijkheid gekregen om enkele vakken van de postgraduale studie *Dutch and Translation* te volgen namelijk *Introduction to Language Technologies*, *Computer Aided Translation* en *Localisation*. Ik vond deze vakken heel belangwekkend en daarom raakte ik in deze problematiek geïnteresseerd. Vorig jaar had ik verder een mogelijkheid om op de vakgroep Engels in Olomouc de vakken *Computer Aided Translation 1* en *Computer Aided Translation 2* te doen. Op zo'n manier heb ik voldoende kennis gekregen om deze scriptie te kunnen schrijven.

Ik hoop dat deze scriptie van nut zal zijn voor studenten Nederlands die in vertalen geïnteresseerd zijn en voor hen een inleiding tot vertaaltechnologieën zal zijn. Het onderwerp is heel breed en interessant, daarom hoop ik dat deze scriptie ook als een motivatie en achtergrond voor verder onderzoek zal dienen.

INHOUDSOPGAVE

<i>Lijst met afkortingen</i>	3
<i>Lijst van afbeeldingen, figuren en tabellen</i>	4
<i>VOORWOORD</i>	7
INHOUDSOPGAVE	8
INLEIDING	11
1 INLEIDING TOT TAALTECHNOLOGIEËN	14
1.1 UITLEG VAN DE BELANGRIJKSTE BEGRIPPEN.....	15
1.2 GESCHIEDENIS EN ONTWIKKELING	16
2 TAAL ALS EEN DYNAMISCH SYSTEEM	18
2.1 TAALTYPOLOGIE	18
2.2.1 <i>Het Tsjechisch</i>	20
2.2.2 <i>Het Nederlands</i>	20
2.2.3 <i>Engels</i>	21
2.2.4 <i>Talenverschillen en daarmee verbonden vertaalproblemen</i>	23
2.3 AMBIGUÏTEIT	23
2.3.1 <i>Lexicale ambiguïteit</i>	24
2.3.2 <i>Idiomen, metaforen, collocations</i>	25
2.3.3 <i>Syntactische ambiguïteit</i>	27
2.3.4 <i>Semantische ambiguïteit</i>	28
2.3.5 <i>Pragmatische ambiguïteit</i>	30
2.3.6 <i>Wereld kennis</i>	31
3 COMPUTERONDERSTEUD VERTALEN	32
3.1 VERTAALGEHEUGEN	33
3.1.1 <i>Principe van vertaalgeheugens</i>	33
3.1.2 <i>Exact match vs. Fuzzy match</i>	34
3.1.3 <i>Alignering</i>	35
3.1.4 <i>Manieren om vertaalgeheugen te verschaffen</i>	36
3.2 VOORDELEN EN NADELEN VAN CAT PROGRAMMA'S	36
3.2.1 <i>Voordelen</i>	36
3.2.2 <i>Nadelen</i>	38
3.3 TERMINOLOGIE	39
3.3.1 <i>Terminologielijst</i>	39
4 VERGELIJKING VAN DE VERTAALPROGRAMMA'S	41
4.1 WORD-BASED VS. STANDALONE (GEBRUIKER INTERFACE)	41
4.2 AANGEBODEN PRODUCTEN	42
4.3 PRIJS.....	45

4.4	SYSTEEMBENODIGDHEDEN	46
4.5	TESTPERIODE (DEMOVERSIE)	47
4.6	ONDERSTEUNDE TALEN	48
4.7	ONDERSTEUNDE BESTANDSFOMATEN	48
4.8	GOOGLE TOOLKIT	49
5	VERDERE TECHNOLOGIEËN EN BEGRIPPEN NOODZAKELIJK VOOR VERTALES VAN VANDAAG.....	52
5.1	OPTISCHE TEKENHERKENNING	52
5.2	LOKALISATIE	53
5.2.1	<i>Waarom noodzakelijk</i>	53
5.2.2	<i>Globalisering / Lokalisatie / Vertalen</i>	56
5.3	MACHINEVERTALING	56
5.3.1	<i>Rule-based MT</i>	58
5.3.2	<i>Corpus-based MT</i>	59
5.4	CROWDSOURCING – EEN SOORT “MODERNE VERTALEN”	60
II	PRAKTISCH GEDEELTE.....	65
6	ENQUÊTE RESULTATEN.....	66
6.1	TALENCOMBINATIES	67
6.2	VERTAALPROGRAMMA’S.....	68
6.2.1	<i>De meest gebruikte programma’s</i>	68
6.2.2	<i>Hoeveel programma’s per vertaler</i>	69
6.2.3	<i>Redenen om een bepaald vertaalprogramma te kiezen</i>	70
6.3	IS PROFESSIONEEL VERTALEN NOG MOGELIJK ZONDER CAT TOOLS.....	71
6.3.1	<i>Vertalers en CAT tools</i>	73
6.3.2	<i>Vertalers NL-CS of/en CS-NL op ProZ.com</i>	74
6.3.3	<i>Gebruik van CAT tools in Tsjechië</i>	75
6.3.4	<i>Voordelen van de programma’s</i>	76
6.3.5	<i>Nadelen van de programma’s</i>	77
6.4	GLOSSARIUM VS. VERTAALGEHEUGEN – WAT BELANGRIJKER	79
6.4.1	<i>Mogelijkheden om glossarium en vertaalgeheugen aan te schaffen</i> ...80	
6.5	COMPETENTIES VAN DE VERTALER VAN VANDAAG.....	81
6.6	OPLEIDING.....	82
6.7	SOORTEN TE VERTALEN TESTEN– SPECIALISATIE VAN VERTALERS	84
6.7.1	<i>Soorten te vertalen teksten</i>	84
6.7.2	<i>Specialisatie</i>	85
6.7.3	<i>Hoe gespecialiseerd</i>	86
6.7.4	<i>Aantal vakgebieden</i>	87
6.7.5	<i>Verschillende soorten te vertalen teksten– nut van CAT tools</i>	88
6.8	LOKALISATIE.....	90
6.8.1	<i>Vertalers en lokalisatie</i>	90
6.9	MT “FRIEND OR FOE?”	91

6.9.1	Zullen vertalers slechts “post-editors” worden?	93
7	CONCLUSIE	95
	<i>Resumé</i>	99
	<i>Summary</i>	100
	<i>ANOTACE</i>	101
	BRONNEN	102
	ELEKTRONISCHE BRONNEN	103
	BIJLAGEN	106
	<i>BIJLAGE I – Enquête voor vertalers</i>	107
	<i>BIJLAGE II – Enquêtes beschikbaar op <http://www.proz.com/polls></i>	110

Inleiding

Toen de Heer neerdaalde om de stad en de toren die de mensen bouwden, te zien, zei Hij: 'Nu zijn ze één volk en spreken zij allen dezelfde taal. Wat zij nu doen is nog maar een begin; later zal geen enkel plan van hen meer te stuiten zijn. Laten Wij neerdalen en verwarring brengen in hun taal, zodat de een niet meer verstaat wat de ander zegt.' En de Heer dreef hen vandaar naar alle kanten de hele aardbodem over, en er kwam een einde aan de bouw van de stad. Daarom noemt men die stad Babel, want de Heer heeft daar verwarring gebracht in de taal van alle mensen, en hen vandaar over de hele aardbodem verspreid.¹

Dit verhaal reflecteert een negatieve visie op het taalverschil, waarbij de differentiatie in feite wordt gezien als een straf van god voor de menselijke overmoed. Sinds de mythische tijden van de Toren van Babel zijn de mensen gedwongen om in verschillende talen te communiceren. De activiteit van tolken en vertalers ontstond om menselijke contacten mogelijk te maken. Na 4000 jaar geschiedenis maken tolken en vertalers nog steeds communicatie mogelijk. Door te vergelijken met machinevertaling kunnen we veel leren over de menselijke vertaling. Sommigen vrezen de concurrentie van de machine voor de mens, maar wat er tot nu toe gebeurd is op het vlak van automatische vertaling, lijkt er veeleer op te wijzen dat mens en machine voor vertaling heel goed complementair kunnen zijn.² Bekend is de uitspraak van de computer-taalkundige en schrijver Hugo Brandt Corstius die zei: *Wat men ook doet, de semantiek gooit roet,*³ zonder wereldkennis zijn bijna alle zinnen ambigu. Als er verschillende interpretaties van een uitdrukking mogelijk zijn, onderscheiden mensen vaak zonder problemen verschillende soorten ambiguïteit en kiezen een goede betekenis. *Deze dubbelzinnigheid die dooreenmens zonder meer dankzij de context wordt weggefilterd, voor de machine soms een onoverkomelijke hinderpaal is.*⁴ Computers kunnen geen foutloze vertalingen produceren maar ze kunnen vertalers wel helpen en hun werk vlugger en efficiënter maken.

De nood aan vertalingen is door de digitale revolutie en vooral door het internet massaal toegenomen. Het aantal documenten dat vertaald moeten worden, groeit elk jaar aanmerkelijk. Om het vertaalproces te versnellen, is gebruik van

¹Genesis 11:1-9 (Willibrordvertaling).

²Willy Vandeweghe, *Duoteksten: Inleiding tot vertaling en vertaalstudie* (Gent: UGent, 2005) 5.

³Vandeweghe 5.

⁴Vandeweghe 171.

verschillende talenttechnologieën noodzakelijk. Het is belangrijk om te beseffen dat deze technologieën geschikt zijn voor repetitieve teksten met een vaste terminologie die letterlijk vertaald moeten worden. Ze zijn dus helemaal *niet* bruikbaar bijvoorbeeld voor literaire vertalingen.

Deze scriptie bevat een theoretisch en een praktisch deel. In het theoretische gedeelte worden alle belangrijke termen en processen uitgelegd die later ook in het praktische gedeelte worden behandeld. Het praktische gedeelte bevat de resultaten van een enquête die werd samengesteld voor vertalers die met CAT-tools werken, namelijk voor vertalers van de taalcombinaties Nederlands-Tsjechisch (hierna: NL-CS), Tsjechisch-Nederlands (hierna: CS-NL), Nederlands-Engels (hierna: NL-EN) en Engels-Nederlands (hierna: EN-NL).

Het eerste hoofdstuk wordt gevormd door een inleiding op de taaltechnologie. Hier worden verschillende taaltechnologieën uitgelegd en behandeld, namelijk computerondersteund vertalen, lokalisatie en machinevertaling. Ook hun verschillen, ontwikkeling, voordelen en nadelen zullen worden besproken.

Het volgende hoofdstuk legt uit waarom computers vertalers niet kunnen vervangen. Talen onderscheiden zich van elkaar en dat maakt het proces van vertalen moeilijker. Hier worden verschillen tussen het Engels, Nederlands en Tsjechisch op grond van hun taaltypologie besproken. Verder worden verschillende types ambiguïteit behandeld, namelijk semantische, lexicale, syntactische en pragmatische ambiguïteit. Computers hebben geen wereldkennis en gevoelens daarom kunnen ze dubbelzinnigheid vaak niet oplossen.

In hoofdstukken drie en vier wordt het computerondersteund vertalen diepgaander besproken. Hier worden de belangrijkste termen uitgelegd, zoals vertaalgeheugen, glossarium, alignement enzv. Verder worden de belangrijkste CAT-tools voorgesteld en vergeleken op grond van hun prijs, functies, ondersteunde talen en bestandsformaten.

In het laatste theoretische hoofdstuk worden nog andere taaltechnologieën behandeld die nauw verbonden zijn met het computerondersteund vertalen en waarvan kennis noodzakelijk is om vandaag de dag vertaler te worden. Het betreft machinevertaling, OCR (optical character recognition), lokalisatie en crowsources.

De basis voor de praktische deel wordt gevormd door een enquête. In het hoofdstuk worden de resultaten van de enquête besproken die gericht was op vertalers NL-CS, CS-NL en NL-EN, EN-NL. De vraagstelling was of er verschillen zijn tussen computerondersteund vertalen van twee kleinere talen en

van een kleinere en een grotere taal. Welke programma's vertalers gebruiken en waarom. Welke opleiding ze hebben afgerond en of ze in een bepaald soort teksten gespecialiseerd zijn. Ik was ook geïnteresseerd in hun opinie over machinevertaling of ze het gebruiken, en of ze ook met lokalisatie van de webpagina's bezig zijn.

Een belangrijke bron van de informatie is de pagina www.ProZ.com. *ProZ.com (de gemeenschap) is een groep professionele taalkundigen, die vertalers, tolken, vertaalbureaus, en hun klanten omvat. ProZ.com (de site) is een markt- en werkplaats: duizenden taalkundigen wisselen hier iedere dag informatie uit over opdrachten en terminologie.*⁵ De gemeenschap ontstond in 1999 en omvatte in 2011 meer dan 500 000 geregistreerde gebruikers in meer dan 200 verschillende landen. Via deze pagina zijn de meeste respondenten voor de enquête gevonden, verder ook informatie over vertaalprogramma's, forums en korte enquêtes die door de gebruikers gemaakt zijn. Hiervan is een deel in deze scriptie gebruikt als aanvullend materiaal.⁶

⁵ *Wat is ProZ.com?* 2012.ProZ. 30 maart 2012
<http://dut.proz.com/faq/general.html#what_is_proz_com>.

⁶ Enquêtes beschikbaar op <<http://www.proz.com/polls>>.

I Een theoretisch gedeelte

1 Inleiding tot taaltechnologieën

1.1 Uitleg van de belangrijkste begrippen

Door het internet en de globalisering is de vraag naar vertalingen steeds groter geworden. Er kunnen twee belangrijke taaltechnologieën worden onderscheiden, namelijk *computerondersteund vertalen* en *machinevertalen*.

Machinevertalen, een calque van het Engelse *machine translation*, is het vertalen van teksten door de computer. Machinevertalen is vooral effectief voor de vertaling van domeinspecifieke teksten zoals technische handleidingen. Het is belangrijk om te beseffen dat de machinevertalingen verre van perfect zijn en dat ze vertalers alleen op enkele gebieden kunnen helpen of vervangen. Vaak moeten de teksten door een vertaler voorbereid (pre-edited) en/of nabewerkt (post-edited) worden.

Vorbewerking betekent dat de taal aangepast wordt zodat die voor de machine begrijpelijker is, bijvoorbeeld alle dubbelzinnige woorden en zinsconstructies worden vervangen.

Nabewerking betekent het verwijderen van ‘automatische’ vertaalfouten en stilistische aanpassing van de tekst. Dat wil zeggen dat bij machinevertalen ervaren linguïsten steeds onvermijdelijk zijn.

Onder niet-professionele vertalers worden op het internet gratis beschikbare vertaalprogramma's, zoals Google Translate of BabelFish steeds populairder. Deze hebben een brede ondersteuning van talen en leveren een ‘ruwe’ vertaling kunnen opleveren. Die vertalingen zijn verre van perfect maar ze worden wel van dag tot dag beter.

Computerondersteund vertalen, overgenomen uit het Engelse *Computer-assisted translation* of *Computer-Aided Translation* wordt afgekort tot CAT. Dit is het vertalen met behulp van computerprogramma's die het vertaalproces ondersteunen en vergemakkelijken. Het is een proces waarbij een menselijke vertaler gebruik maakt van software om grotere precisie en efficiëntie te bereiken. Het is belangrijk om het verschil tussen machinevertalen en computerondersteund vertalen duidelijk te zien. Bij machinevertalen is de computer degene die de vertaling verricht. Bij computerondersteund vertalen maakt de vertaler zelf de vertaling en maakt hierbij gebruik van computerprogramma's.

Met computerondersteund vertalen is *lokalisatie* (localisation) nauw verbonden. Lokalisatie betekent dat de documentatie wordt aangepast aan de eisen van de markt van de doeltaal. Het proces omvat méér dan alleen vertalen. Het

vertalen is slechts een onderdeel van een gecompliceerd proces van adaptatie van een product aan de doelmarkten (zie hoofdstuk 5.2).

1.2 Geschiedenis en ontwikkeling

Sinds het ontstaan van de computer is men bezig met het ontwikkelen van vertaalsystemen. In het begin was de bedoeling om automatische vertaalsystemen te ontwikkelen die in plaats van de vertalers zouden werken. Een van de eerste experimenten op het gebied van automatisch vertalen was het vertaalsysteem Russisch-Engels tijdens Koude Oorlog. Het werd in 1954 in de Verenigde Staten aan de universiteit van Georgetown ontwikkeld. Het was een zogenaamd direct systeem met behulp van een elektronisch woordenboek van 250 woorden en een beperkte grammatica. Er werden woord-voor-woord-vertalingen tot stand gebracht.⁷

Een van de succesvolle projecten was bijvoorbeeld TAUM-Météo. Het was een automatisch vertaalsysteem voor weerberichten (EN>FR), ontwikkeld aan de universiteit van Montreal. Het succes van dit project lag aan de steeds terugkerende en stereotiepe woordenschat.

Een pessimistischrapport over MT:

Er kwam een einde aan het enthousiasme en vooral aan de financiële subsidies met een rapport van de Automatic Language Processing Advisory Committee (ALPAC) in 1966. Deze commissie betoogde dat er geen tekort aan menselijke vertalers bestond en dat er bovendien op korte termijn geen bruikbare vertalingen van wetenschappelijke teksten te verwachten waren. Op basis van het ALPAC-rapport beëindigde de overheid van de Verenigde Staten de subsidiëring van machinevertalingen.⁸ Automatisch vertalen is volgens de commissie niet alleen duurder en tijdrovender dan gewoon vertalen, het zal ook nooit de kwaliteit ervan kunnen evenaren. De hoofdredenen daarvoor zijn het gebrek aan interpretatievermogen van een machine die op cognitief en contextueel gebied slecht scoort. Het rapport betekende in de Verenigde Staten een stopzetting van de financiering voor onderzoek gericht op automatischvertalen, maar in Europa, Canada en Japan ging men door. Men begon zich meer te concentreren op

⁷Vandeweghe 172-5.

⁸H.A.Becker, *Generaties en de strijd tegen taalbarrières* (Utrecht: Universiteit Utrecht, 2001), 18.

systemen waarbij mens en machine samenwerken, met andere woorden waar dus sprake is van computerondersteund vertalen.⁹

Automatisch vertalen wordt toch veel gebruikt. Met name in zogenaamde beperkte domeinen, zoals weerberichten of technische handleidingen, waarin maar een beperkt aantal begrippen gebruikt wordt,” legt Tilburgse taaldeskundige en hoogleraar Antal van den Bosch uit. *“Helemaal overbodig zullen menselijke vertalers nooit worden. Als de tekst een emotie overbrengt, zoals een gedicht of een roman, zal een mens het altijd beter kunnen vertalen dan een computer.”*¹⁰ Het betekent dat de vertalers de taaltechnologieën als een concurrentie niet hoeven te vrezen. Bij sommige soorten teksten kunnen vertalers nieuwe technologieën ten eigen behoeve benutten en bijvoorbeeld bij literaire teksten zijn de programma's feitelijk onbruikbaar.

⁹Vandeweghe172-5.

¹⁰Erica Renckens. 23 februari 2009. “I sit with the hands in the hair”. Kennislink. 8 februari 2012 <<http://www.kennislink.nl/publicaties/i-sit-with-the-hands-in-the-hair>>.

2 Taal als een dynamisch systeem

Waarom kunnen vertaalprogramma's geen vertalingen van hoge kwaliteit leveren? *Any attempt to replace Human Translation totally by machine translation would certainly face failure ... The automatic translation ... has proved its weakness, most of the time, when compared with a human translation. The human translator is the only subject in a position to understand the different cultural, linguistic and semantic factors contributing to leaving the same effect that is left in the source text, in the target text.*¹¹ Elke taal vormt een uniek systeem met eigen eigenschappen wat betekent dat talen ten opzichte van elkaar verschillen. Dat kan echter moeilijkheden veroorzaken tijdens het proces van vertalen ook voor de mensen maar vooral voor computers die geen menselijke kennis hebben. In dit hoofdstuk zullen de talen (Nederlands, Engels, Tsjechisch) ten eerste op de basis van hun taaltypologie besproken worden. Ten tweede zullen verschillende soorten ambiguïteit behandeld worden.

2.1 Taaltypologie

Talen kunnen op grond van hun morfologie in vier basistypes onderverdeeld worden. Er zijn isolerende (analytische), agglutinerende, synthetische (flecterende), en polysynthetische talen.¹²

Het meest extreme geval van een *analytische taal* is een isolerende taal. In een isolerende taal bestaan geen of weinig suffixen. In die talen kunnen de meeste morfemen zelfstandig zijn. Grammaticale relaties worden binnen analytische talen uitgedrukt door de woordvolgorde, intonatie en klemtoon. Tot dit type behoren Chinees en Vietnamees maar ook het Engels en Nederlands.

Een *agglutinerende taal* is een taal waarin woorden in morfemen gesplitst kunnen worden. Er zijn duidelijke morfeemgrenzen, en elke grammaticale betekenis wordt uitgedrukt door een afzonderlijk morfeem. Een voorbeeld hiervan is het Japans of het Hongaars.

¹¹Abdelhak Jebbar. 2012. *Machine Translation vs human Translation*. 5 april 2012
<<http://www.translationdirectory.com/articles/article1326.php>>.

¹²Masayoshi Shibatani, Theodora Bynon, *Approaches to language typology* (Oxford: Oxford University Press, 1995) 6.

In een *synthetische taal* zijn de grenzen tussen de morfemen minder duidelijk, en morfemen kunnen meer dan één grammaticale betekenis tegelijk uitdrukken. In fleterende talen wordt gebruik gemaakt van vervoegingen, verbuigingen en naamvallen. Tot dit type behoren bijvoorbeeld het Latijn maar ook het Tsjechisch.

In een *polysynthetische taal* worden morfemen samengevoegd om lange samengestelde woorden te vormen die als een volledige zin dienst kunnen doen. Een voorbeeld hiervan zijn bijvoorbeeld Eskimo talen.¹³

Deze vier taaltypen zijn slechts theoretische concepten om talen te onderscheiden. In de praktijk kunnen talen niet echt worden verdeeld in types. De meeste talen hebben kenmerken van meer dan één type, bijvoorbeeld hoewel het Engels bijna volledig een analytische taal is, heeft deze taal ook een paar kenmerken van een synthetischetaal, bijvoorbeeld bij vervoeging van werkwoorden in de derde persoon enkelvoud (*he goes*) of genitief van de persoonsnamen (*Peter's house*).

2.1.1 Het Tsjechisch¹⁴

Het Tsjechisch is een West-Slavische taal die ongeveer door 9,5 miljoen personen in Tsjechië wordt gesproken. Er zijn ook Tsjechische gemeenschappen in de buurlanden Oostenrijk, Polen, Slowakije en Duitsland, maar ook in de Verenigde Staten en Canada.

Het Tsjechisch is een synthetische taal die een rijk verbuigings- en vervoegingssysteem heeft. De zelfstandige naamwoorden hebben drie geslachten – vrouwelijk, mannelijk (levend, niet-levend) en onzijdig. Het Tsjechisch heeft zeven naamvallen (nominatief, genitief, datief, accusatief, vocatief, locatief en instrumentalis). De zelfstandige naamwoorden, bijvoeglijke naamwoorden, voornaamwoorden en telwoorden krijgen hierbij volgens hun rol in de zin verschillende uitgangen. De werkwoorden worden vervoegd. Soms veranderen naast verschillende uitgangen ook medeklinkers en klinkers in de woordkern. Verder hebben morfemen vaak meer dan één grammaticale betekenis (zie voorbeeld 1):

¹³Jaroslav Popela, Vykypěl Bohumil, Boček Vít, *Skaličkova jazyková typologie* (Brno: Masarykova univerzita, 2006) 9-12.

¹⁴Bernard Comrie, *The Slavonic Languages* (Routledge: London, 2002) 455-513.

- (1) De uitgang *-l* van de persoonsvorm *dělal* (hij deed) heeft vier grammaticale betekenissen: derde persoon, enkelvoud, mannelijk geslacht en verleden tijd.

Persoonlijke voornaamwoorden worden alleen als onderwerp uitgedrukt als de nadruk op de persoon ligt of als het wegens onduidelijkheid nodig is. De taal bevat verder een aantal specifieke diakritische tekens, namelijk het “omgekeerd dakje” (*háček*, ˇ), het accentteken (*čárka*, ´) en het rondje/het kringetje (*kroužek*, °). De klemtoon ligt voornamelijk op de eerste lettergreep. In tegenstelling tot het Engels heeft het Tsjechisch een grote vrijheid in de zinsbouw en gebruikt het geen lidwoorden.

2.1.2 Het Nederlands

Het Nederlands is een Germaanse taal, die wereldwijd door ruim 25 miljoen mensen wordt gesproken. Het Nederlands is de officiële taal van Nederland, België, Suriname, Aruba, Curaçao en Sint-Maarten. Hoewel het Nederlands, Duits en Engels als Germaanse talen een gemeenschappelijke voorvader hebben, is er een opvallend verschil tussen Nederlands en Engels aan de ene kant en Duits aan de andere kant, namelijk de naamvallen. Het Nederlands is een isolerende taal die bepaalde kenmerken van een synthetische taal heeft. Er zijn minder woordverbuigingen dan in het verwante Duits. In het Nederlands worden werkwoorden wel vervoegd, maar de vervoeging is niet zo complex als in het geval van flecterende talen.¹⁵

Het morfologische systeem van het Nederlands is niet zo uitgebreid als bijvoorbeeld de Duitse of Tsjechische. De belangrijkste verbuigingen zijn de meervoudsvorm (2), de verbuigings *e* van het bijvoeglijk naamwoord (3) en het diminutief (4) van de zelfstandige naamwoorden:¹⁶

(2) stoel – stoelen, lente – lentes, auto – auto’s, kind – kinderen

(3) een *klein* kind, *kleine* kinderen, een *kleine* man

(4) klas – klasje, kan – kannetje, boom – boompje, been – beentje

Een voorbeeld van vervoeging zijn bijvoorbeeld de werkwoordsvormen van de tegenwoordige tijd van de werkwoorden, zie het werkwoord *gaan* (5):

¹⁵Jos Thijssens, Eric Tytgat, Bart Vandenberghe, *Taalkracht tien 4* (Malle: De Sikkel, 1996) 452.

¹⁶“Vebuiging“. 2012. Nederlandse Taalunie, Taaladvies.net. 9 februari 2012 <http://woordenlijst.org/leidraad/lijst_van_vaktermen/verbuiging/>.

(5) Ik *ga* / jij *gaat* / hij, zij, het, *ugaat* / wij, jullie, zij *gaan*

In het Nederlands is het gebruik van naamvallen grotendeels verdwenen. Slechts bij de persoonlijke voornaamwoorden (6) is het onderscheid tussen verbogen en onverbogen vormen in de meeste gevallen behouden:

(6) Ik - mij, jij – jou, hij – hem, zij – haar, wij – ons, zij –hen / hun

Het enige verschil tussen accusatief (7) en datief (8) in het moderne Nederlands zijn vormen *hen* – *hun* van de persoonlijke voornaamwoorden.

(7) Ik zie hen.

(8) Ik geef hun het glas.

In de praktijk worden deze vormen echter veelal door elkaar gebruikt.

Bij de andere woordsoorten zijn de oude naamvals vormen alleen in verstaende uitdrukkingen bewaard gebleven. Dat wil zeggen dat naamvallen in de hedendaagse Nederlandse standaardtaal vrijwel niet meer gebruikt worden, behalve in vaste combinaties (9):

(9) ‘s morgens, in groten getale, het Koninkrijk der Nederlanden etc.

Het Nederlands is in zijn moderne vorm overwegend een analytische taal.¹⁷ De functie die een woord in de zin heeft, is nu te zien aan de voorzetsels (10):

(10) *van* mij, *naar* huis, *voor* mijn vriend enz.

2.1.3 Engels

Het Engels behoort tot de Germaanse talen. Modern Engels, soms beschreven als de eerste wereldwijde *lingua franca*,¹⁸ is de belangrijkste internationale taal voor radio en televisie, communicatie, wetenschap, luchtvaart, toerisme, de zakenwereld en andere.¹⁹ Er zijn 328 miljoen moedertaalsprekers van Engels ter wereld. Wat betreft aantal moedertaalsprekers is Engels de derde grootste taal na Chinees en Spaans. De taal wordt in totaal in 112 landen gesproken. Wat betreft het aantal

¹⁷“Naamvallen“. 2009. Nederlandse Taalunie, Taaladvies.net. 30 maart 2012 <<http://taaladvies.net/taal/advies/tekst/30/naamvallen/>>.

¹⁸De term *lingua franca* verwijst naar een taal die op grote schaal als gemeenschappelijk communicatiemiddel wordt gebruikt tussen mensen met een verschillende moedertaal.

¹⁹David Graddol, *The Future of English: A guide to forecasting the popularity of the English language in the 21st century* (The British Council, 2000), 3-5.

landen zit Engels echter op de eerste plaats ter wereld. Ongeveer 600 miljoen mensen spreken verder Engels als eerste vreemde taal²⁰.

Een van de redenen waarom het Engels zo'n wereldverspreide taal werd, is wellicht de structuur van de taal. Engels is een analytische taal die weinig verbuigingen en vervoegingen kent. Een voorbeeld van vervoeging van werkwoorden is de derde persoon enkelvoud (11) waar de persoonsvorm in de tegenwoordige tijd de uitgang *-s* krijgt:

(11) *I/you/we/they play– he/she/it plays*

Net als in het Nederlands is het gebruik van naamvallen bijna volledig verdwenen en zelfs in nog grotere mate. Alleen de genitief (12) van de persoonsnamen wordt bewaard. De genitief wordt met 's geschreven maar het is ook mogelijkheid om die analytisch om te schrijven met het voorzetsel *of*:

(12) *John's house – a house of John*

Uit het bovenstaande voorbeeld blijkt dat net als in het Nederlands de voorzetsels bepalen welke functie een woord in de zin heeft.

Het Engels heeft een vaste SVO-woordvolgorde (onderwerp – werkwoord – voorwerp). In het Tsjechisch wordt er door wijzigingen van de woordvolgorde andere nadruk op het gewenste zinsdeel gegeven. Om een zinsdeel te benadrukken is er in het Engels echter een speciale zinsconstructie noodzakelijk. Als goede voorbeeld kunnen onderstaande zinnen dienen (13) *there is* constructie en (14) een passieve zin:

(13) *There is a pen on the table. – Na stole je pero.*

(14) *The house was built by Peter. – Dům postavil Petr.*

Het benadrukken van de zinsdelen is nauw verbonden met *thema* en *rhema* verdeling binnen een zin. Het thema staat over het algemeen vooraan in de zin en wordt gevolgd door het rhema (in cursief – zie *pen* en *Peter*), dat hoofdzakelijk extra en nieuwe informatie toevoegt met betrekking tot het thema dat van de voorafgaande context al bekend is en betekent dus een oude informatie.

²⁰M. Paul Lewis. 2009. *Ethnologue: Languages of the World*. 13 februari 2012
<http://www.ethnologue.com/ethno_docs/distribution.asp?by=size>.

2.1.4 Talenverschillen en daarmee verbonden vertaalproblemen

Hoewel Engels en Nederlands beide Germaanse en analytische talen zijn, verschillen ze van elkaar. Het Tsjechisch is bovendien een taal die tot een andere talengroep behoort. Het is een Slavische taal die synthetisch is. Dat betekent dat er in vergelijking met Engels en Nederlands een rijk morfologiesysteem is. Verder zijn er ook verschillen in de zinsbouw – de woordvolgorde is in het Tsjechisch behoorlijk vrij maar in het Engels en Nederlands ligt die meer vast. In het Tsjechisch ontbreken er verder lidwoorden. Alle taalverschillen maken woord-voor-woord vertaling onmogelijk en daarom ook het vertaalproces moeilijker.

2.2 *Ambigüiteit*

*Ambiguity is one of the bottle-necks of Machine Translation, and of any computer program that tries to understand human language. Many linguistic phenomena result in some kind of ambiguity: words, for example, may belong to more than one syntactic category (syntactic homonymy), and the program must be able to decide which is the correct one ... In an MT application we find some additional problems. First, MT systems make little use of extrasentential context to resolve cases of ambiguity due to the complexity involved in implementing this type of strategy. Rather, most systems work on a sentence-by-sentence basis, making some sort of best-guess according to different heuristics.*²¹

Ambigüiteit is een van de karakteristieke eigenschappen van natuurlijke taal. Woorden, uitdrukkingen en zinnen hebben vaak meer dan één vaste betekenis en de juiste betekenis hangt in zulke gevallen van de context af. Voor mensen is ambigüiteit vrijwel nooit een probleem omdat ze over complexe cognitieve mechanismen beschikken die voor correcte interpretatie zorgen. Communicatie tussen mens en computer kan echter niet ambigu zijn, omdat computers in enen en nullen ‘denken’. Om succesvol met een computerprogramma te kunnen communiceren is het noodzakelijk om ambigüiteit te vermijden door volledig expliciete uitingen.

Ambigüiteiten kunnen verschillende oorzaken hebben. *Lexicale* ambigüiteit wordt veroorzaakt door een woord of woordgroep die meerdere betekenissen heeft. *Syntactische* ambigüiteit betekent dat een uiting syntactisch op verschillende manieren geanalyseerd kan worden. Naast lexicale en syntactische

²¹Gloria Álvarez Benito. 1994. *Contextual Deletion of Object and Ambiguity in Machine Translation*. 25 maart 2012. <http://rua.ua.es/dspace/bitstream/10045/6012/1/RAEI_07_02.pdf>.

analyse noemen Arnolds et al. drie soorten kennis die noodzakelijk zijn om ambiguïteit te herkennen: *semantische kennis*, d.w.z. linguïstische kennis die context-onafhankelijk is, *pragmatische kennis*, d.w.z. linguïstische kennis die context-afhankelijk is; en *wereldkennis*, d.w.z. algemene, non-linguïstische kennis die nauw verbonden is met de pragmatische kennis.²² Een computer zou in het geval van ambiguïteit meer dan één betekenis van de zin interpreteren die mensen vaak helemaal niet zouden merken, bijvoorbeeld het woord *bank* – het is voor een mens duidelijk of er een financieel instituut of een stoel bedoeld wordt. Bij het interpreteren van uitingen speelt vooral context een belangrijke rol.

2.2.1 Lexicale ambiguïteit

Er bestaan twee types *lexicale* ambiguïteit, namelijk *syntactische* en *semantische*. Syntactische lexicale ambiguïteit is ambiguïteit van een woordsoort in het geval dat één woord tot twee of meer verschillende woordsoorten behoort bijvoorbeeld zelfstandig naamwoord tegenover werkwoord (15), bezittelijk voornaamwoord versus werkwoord (16) of bijvoeglijk naamwoord versus voorzetsel(17):²³

(15) *sla*: een bladgroente, zelfstandig naamwoord
sla: “geef een klap, werkwoord

(16) *zijn*: “van hem“, bezittelijk voornaamwoord
zijn: “wezen“, werkwoord

(17) *naar*: “onaangenaam“, bijvoeglijk naamwoord
naar: “in de richting van“, voorzetsel

Semantische lexicale ambiguïteit wordt veroorzaakt door woorden die tot dezelfde woordgroep behoren maar meer dan één betekenis hebben (homonymie en polysemie). *Homoniemen* zijn verschillende woorden van dezelfde woordsoort die wel dezelfde vorm hebben, maar verder niets met elkaar te maken hebben. Woorden wijken van elkaar af op grond van hun etymologie en semantiek. Het woord *ezel* kan verwijzen naar een dier (en naar personen die daar een zekere overeenkomst mee vertonen) of naar een stuk schildersgereedschap. Wanneer we een zin als *De ezel stond in het veld* in isolatie proberen te interpreteren valt niet te

²²Arnold Douglas, Lorna Balkan, Siety Meijer, R. Lee Humphreys, Louisa Sadler, *Machine Translation: An Introductory Guide* (London: NCC Blackwell, 1994) 130-139.

²³Geert Adriaens, Steven Lawrence Small, Garrison Weeks Cottrell, Michael K. Tanenhaus, *Lexical ambiguity resolution: Perspectives from Psycholinguistics, Neuropsychology, and Artificial Intelligence* (CA: Morgan Kaufman Publishers, 1988) 5.

beslissen welke betekenis bedoeld wordt.²⁴ Maar vaak kan uit de context worden afgeleid wat de betekenis is van het woord. Het meest bekende voorbeeld is waarschijnlijk het woord *bank* (18):

(18) Ze ging zitten op de *bank*.

Hij bracht zijn spaargeld naar de *bank*.

Bij het eerste voorbeeld van *bank* gaat het om een geldinstelling, bij het tweede voorbeeld gaat het om een zitmeubel. Bij dit voorbeeld is het duidelijk dat de beide woorden geen relatie hebben qua betekenis.

Bij *polysemie* gaat het ook om een woord, die op verschillende manieren gebruikt kan worden. Het verschil hierbij is dat de betekenissen van het woord dat op verschillende manieren gebruikt is, wèl aan elkaar verwant zijn. Denk aan een *blad* (19):

(19) een *tafelblad*

een *blad* van een plant

een *dienblad*

een *ochtendblad*

Deze woorden zijn allemaal afgeleid van een *blad* van een plant: iets dat dun, plat en breed is. Dat wordt gezien als voldoende gelijkenis om de woorden als verwant aan elkaar te zien. Ze hebben immers dezelfde ontstaanswijze, dus ze zijn polyseem.²⁵

2.2.2 Idiomen, metaforen, collocaties

Hutchins noemt volgende drie problemen die in het proces van het vertalen van specifieke woorden kunnen ontstaan, namelijk collocaties, idiomen en metaforen.²⁶ *These linguistic devices, in a manner similar to lexical ambiguity, present a comprehender with a necessity of selecting from among a set of possible interpretations. Presumably, a person uses context to help this selection process, as one does with ambiguous word meanings.*²⁷ Begrip voor idiomen is, net als

²⁴Gosse Bouma and Gertjan van Noord. 1998. "Natuurlijke-taalinterfaces". Vakgroep Alfa-informatica & BCN Rijksuniversiteit Groningen. 14 februari 2012 <<http://www.let.rug.nl/vannoord/papers/breinmakers/nli.html>>.

²⁵*Homonymie, polysemie en synonymie*. 30 oktober 2011. Kunst en cultuur 25 maart 2012 <<http://kunst-en-cultuur.infonu.nl/taal/85815-homonymie-polysemie-en-synonymie.html>>.

²⁶Hutchins, John. 1991. "Why computers do not translate better." 5-6. 5 april 2012 <<http://www.mt-archive.info/Aslib-1991-Hutchins.pdf>>.

²⁷Lexical ambiguity resolution 277.

voor metaforen en collocaties, afhankelijk van cultuur en achtergrond. Dat wil zeggen dat ze taalspecifiek zijn. Ze moeten aangeleerd worden en een leterlijk vertaling leidt vaak tot een foutieve en niet idiomatische tekst.

Idioom

*Idioom is een combinatie van woorden die iets anders betekent dan de losse woorden samen.*²⁸ Het is een taaluiting die je niet kunt vertalen, het heeft een vaste, ondoorzichtige (idiosyncratische) structuur, wat betekent dat woord-voor-woord vertaling onmogelijk is. Idiomen kun je niet encoderen, maar ook niet decoderen, omdat ze een niet-compositionele betekenis hebben. Voorbeelden van verschillen:

Tabel 1 Idiomen

Nederlands	Engels	Tsjechisch
Als een olifant in een porseleinkast	Like a bull in a china shop	Jako slon v porcelánu
De pijp aan Maarten geven	To kick the bucket	Natáhnout bačkory

Collocatie

*Collocatie is een idiomatische verbinding van woorden.*²⁹ Het verschil tussen idiomen en collocaties is als volgt: collocaties kun je wel decoderen als je ze niet kent, maar niet encoderen als je ze niet kent. Idiomen kunt je noch encoderen noch decoderen. Enkele voorbeelden (20):

(20) *zware roker* maar niet **zware eter*
bloedheet maar niet **bloedkoud*
vrede stichten maar niet **vrede maken*
bang voor maar niet **jaloers voor*

Deze voorbeelden suggereren wat collocaties zijn: vaste verbindingen met een zekere mate van onvoorspelbaarheid in de keuze van de gebruikte woorden.

Metafoor

*Metafoor is een vorm van beeldspraak, waarbij de afbeelding of beschrijving berust op een vergelijking.*³⁰ Sommige metaforen kunnen zijn gelijk in het Neder-

²⁸*Idioom*. 2012. Encyclo. 15 maart 2012 <<http://www.encyclo.nl/begrip/idioom>>.

²⁹“Collocatie“. Encyclo. 27 maart 2012 <<http://www.encyclo.nl/zoek.php?woord=collocatie>>.

³⁰“Metafoor“. Encyclo. 27 maart 2012 <<http://www.encyclo.nl/begrip/Metafoor>>.

lands, Engels en ook Tsjechisch. Er zijn voorbeelden van enkele gelijke dieren-metaforen:

Tabel 2 *Gelijke metaforen*

Nederlands	Engels	Tsjechisch
Zo sterk als een beer	As strong as s bear	Silný jako medvěd
Zo trots als een pauw	As proud as s peacock	Pyšný jako páv
Zo stil a een muis	As quiet as a mouse	Tichý jako myška
Zo vies als een varken	As dirty as s pig	Líný jako prase

Maar met het vertalen van metaforen moeten we voorzichtig zijn. Er zijn veel gevallen die van elkaar afwijken als in deze gevallen taal- en cultuur specifieke referenten gebruikt worden:

Tabel 3 *Afwijkende metaforen*

Nederlands	Engels	Tsjechisch
Zo doof als een kwartel	As deaf as a post	Hluchý jako pařez
Zo dronken als een tor	As drunk as a lord	Opilý jako Dán
Zo oud als de weg naar Rome	As old as a hills	Starý jako černé uhří
Zo nijdig als een spin	As angry as a spider	Naštvaný jako býk

2.2.3 Syntactische ambiguïteit

Sommige zinnen kunnen syntactisch (structureel) dubbelzinnig zijn. Zo'n ambiguïteit wordt veroorzaakt door de mogelijkheid van twee of meer verschillende grammaticale ontledingen van een zin. Dit betekent dat een zin op meerdere manieren syntactisch geanalyseerd kan worden wat tot verschillen in betekenis kan leiden. Voor een computer is het erg lastig te bepalen welke van de twee interpretaties in de bepaalde zin bedoeld wordt. Om de ambiguïteit op te lossen is wereldkennis noodzakelijk (21):

- (21) *Hij at een pizza zonder olijven.* (de olijven horen bij een pizza)
Hij at een pizza zonder een mes. (mes is een instrument)
Hij at een pizza zonder Piet. (Piet is afwezig)

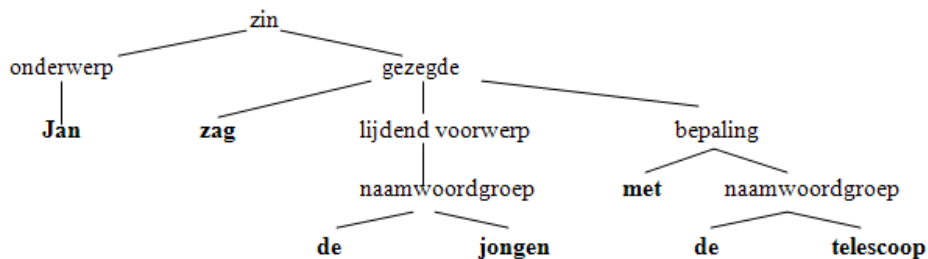
Er bestaan echter ook zinnen die ook voor mensen verwarrend kunnen zijn. In zo'n geval moet er om een context of om aanvullende informatie gevraagd worden (22):

- (22) *Jan zag de jongen met de telescoop.*

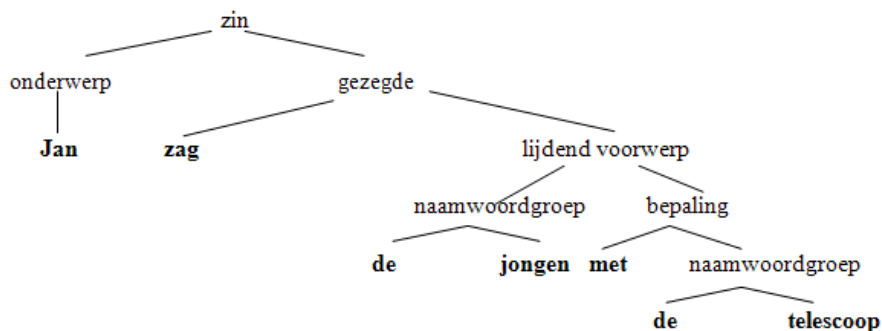
De betekenis van de bovenstaande zin kan worden achterhaald door middel van syntactische analyse, namelijk via een analyseboom. Het is noodzakelijk om de zin in woordgroepen te onderscheiden en hun onderlinge relatie vast te leggen.

Het is duidelijk dat *Jan* degene is die ziet, terwijl *de jongen* de geziene is. Dat wil zeggen dat *Jan* onderwerp is van gezegde, terwijl *de jongen* direct object is. Twee mogelijke analysebomen voor deze zin zijn de volgende:

Afbeelding 1 Syntactische boom I



Afbeelding 2 Syntactische boom II



De eerste analyseboom kan worden gebruikt om de lezing af te leiden waarbij de telescoop een instrument van Jan is. In de tweede analyseboom kan de lezing worden afgeleid waarbij de jongen toevallig de telescoop bij zich heeft, maar waarbij ongenoemd blijft hoe Jan de jongen heeft gezien, bijvoorbeeld met de bril, de telescoop of anders. Deze voorbeelden tonen aan waarom syntactische analyse van belang kan zijn om de betekenis van een zin te begrijpen.³¹

2.2.4 Semantische ambiguïteit

*Factual (semantic) knowledge is context-independent and shared, in more or less detail, across a cultural community.*³²

³¹Gosse Bouma and Gertjan van Noord. 1998. "Natuurlijke-taalinterfaces". Vakgroep Alfa-informatica & BCN Rijksuniversiteit Groningen. 14 februari 2012 <<http://www.let.rug.nl/vannoord/papers/breinmakers/nli.html>>.

³²Richard J S Wise. 2012. "Localising semantic memory with functional neuroimaging." British Medical bulletin. 5 april 2012. <<http://bmb.oxfordjournals.org/content/65/1/95.full>>.

Semantische kennis is een linguïstische kennis die context onafhankelijk is. In sommige gevallen is semantische kennis noodzakelijk om ambiguïteit te vermijden. Een nuttige methode is het gebruik van *semantische kenmerken* (semantic features). Semantische kenmerken drukken enkele semantische eigenschappen van het woord uit. Een klassiek voorbeeld daarvan is:

Man [+HUMAN], [+MALE], [+ADULT]

Woman [+HUMAN], [-MALE], [+ADULT]

Boy [+HUMAN], [+MALE], [-ADULT]

Girl [+HUMAN], [-MALE], [-ADULT]³³

Semantische kenmerken kunnen ons helpen bij zinsontleding. Een voorbeeld zin van Hutchins (23):³⁴

(23) *The monkey ate the banana because it was hungry.*

Bij het vertalen naar het Tsjechisch moeten twee dubbelzinnigheden opgelost worden: het geslacht van het bijvoeglijke naamwoord *hungry* en de geschikte vertaling van het werkwoord *ate*.

Money [-HUMAN], [+ANIMATE], [-MALE]

Banana [-HUMAN], [-ANIMATE], [+MALE]

Honger is verbonden met een levend wezen in dit geval dus de *monkey* en niet de *banana*. *Monkey* is in het Tsjechisch vrouwelijk, er is dus nog een goede uitgang nodig (*hladov-á*). In het Tsjechisch is er bovendien nog een verschil bij het werkwoord *ate* (*jedla x žrala*) voor human / non-human semantische eigenschappen. Zo'n analyse is noodzakelijk om de kwaliteit van machinevertalen te verbeteren.

³³Hutchins, John. 1991. "Why computers do not translate better." 5-6. 5 april 2012 <<http://www.mt-archive.info/Aslib-1991-Hutchins.pdf>>.

³⁴Petr Mikulík, *A Study of selected Machine Translation Systems*. Diplomová práce. (Brno: Masaryk Universiteit, 2009) 8.

2.2.5 Pragmatische ambiguïteit

Volgens Arnolds et al. is de meest voorkomende type van context gebaseerde ambiguïteit *deixis* (anaphora resolution). Deiktische elementen zijn elementen in een tekst die verwijzen naar de uitingssituatie.³⁵ We kunnen weer de zin: *The monkey ate the banana because it was hungry* gebruiken. De vertaler moet beslissen of het voornaamwoord *it* naar *monkey* of naar *banana* verwijst. Dit probleem kan weer opgelost worden door semantische analyse. Een levend wezen kan honger hebben dus *monkey* en niet *banana*. In het Tsjechisch is *monkey* vrouwelijk en *banana* mannelijk daarom is het noodzakelijk om te beslissen welk woord het antecedent is.

Persoonsdeixis: verwijst naar een persoon in de uitingscontext. Er wordt verwezen naar de rolverdeling binnen een gesprek. Er zijn drie mogelijk deelnemers: de spreker, de hoorder en mensen die geen hoorder en geen spreker zijn (ik, hij, hem, hun, enz.).

Tijdsdeixis: wordt er verwezen naar tijd in verhouding tot de tijd van het gesprek. Ook de tijden van het werkwoord zorgen voor een deiktische verankering van wat men zegt in de uitingssituatie (gisteren, vandaag, morgen, vorige week, zondag enz.)

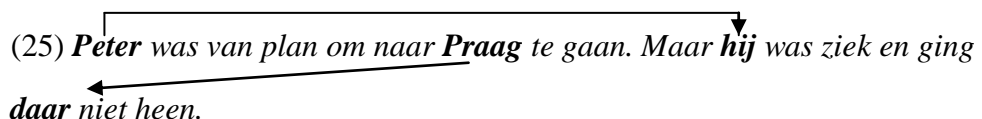
Plaatsdeixis: er wordt verwezen naar de locatie in verhouding tot de mensen die deelnemen aan het gesprek (hier, daar, op die plaats, enz.):³⁶

(24) *Ik zag hem daar gisteren.*

In de bovenstaande zin (24) zien we een voorbeeld van *persoonsdeixis* (ik, hem), *tijdsdeixis* (gisteren, imperfectum) en *plaatsdeixis* (daar). Om deze zin volledig te begrijpen hebben we een context nodig.

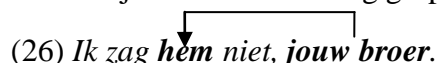
Anafoor: verwijst naar iets wat eerder in de tekst gespecificeerd werd (25):

(25) *Peter* was van plan om naar *Praag* te gaan. Maar *hij* was ziek en ging *daar* niet heen.



Katafoor: verwijst naar iets wat nog gespecificeerd moet worden (26):

(26) *Ik zag hem* niet, *jouw broer*.



³⁵Arnold Douglas, Lorna Balkan, Siety Meijer, R. Lee Humphreys, Louisa Sadler, *Machine Translation: An Introductory* (Guide. London: NCC Blackwell, 1994) 130-139.

³⁶*Pragmatiek*. 2012. Introduction to Dutch Linguistics. 40 april 2012 <http://www.dutch.ac.uk/an/SP_LINKS_UCL_POPUP/SPs_dutch/linguistics/meaning_pragmatics.html>.

Een menselijke vertaler gaat in zulke gevallen zijn/haar pragmatische kennis benutten. Een context helpt om te beslissen welke mogelijkheid juist is. Voor de computerprogramma's betekent het vaak een hinderpaal en een potentiële fout.

2.2.6 Wereld kennis

*Natural languages are rich in ambiguities of many kinds. Furthermore, many sentences that do not seem ambiguous to humans, due to their extensive world knowledge, may present ambiguities to machine translation systems.*³⁷ Wereldkennis is algemene kennis die computers ontbreken en niet kunnen aanleren. *Het feit dat mensen meestal geen enkele moeite hebben om aan zinnen een eenduidige interpretatie te geven heeft te maken met semantische kennis, maar bovenal met extra-linguïstische kennis, dat wil zeggen kennis van hoe de wereld in elkaar zit, kennis van het domein, kennis van hoe mensen met taal omgaan, etc. Tot nu toe is het onmogelijk om dergelijke kennis volledig te formaliseren.*³⁸

Een goed voorbeeld is een uitdrukking *on the river*³⁹ (27) die met verschillende zelfstandige naamwoorden een andere betekenis krijgt:

- (27) *a house on the river* (een huis staat aan de oever)
a boat on the river (een boot aan het water)

Meer voorbeelden (28), (29):⁴⁰

- (28) Put *the paper in the printer*. Then switch **it** on.

=> mensen begrijpen dat *it* verwijst naar een printer die degene is die ingeschakeld kan worden.

- (29) *Mary hit her bag against the vase*, **it** broke.

=> mensen begrijpen dat *it* verwijst naar de *vase* omdat die breekbaar is en kapot kan gaan.

³⁷Pascual Cantos Gómez, *Lexical ambiguity, dictionaries and corpora* (Universidad de Murcia, 1996) 16.

³⁸Nijholt, A. 1994. "Parlevink in context: meer dan language engineering". 25 maart 2012 <<http://doc.utwente.nl/67128/1/eniac.pdf>>.

³⁹P. G. J. van Sterkenburg, *Unity and diversity of languages* (Amsterdam: John Benjamins Publishing Co, 2008) 198.

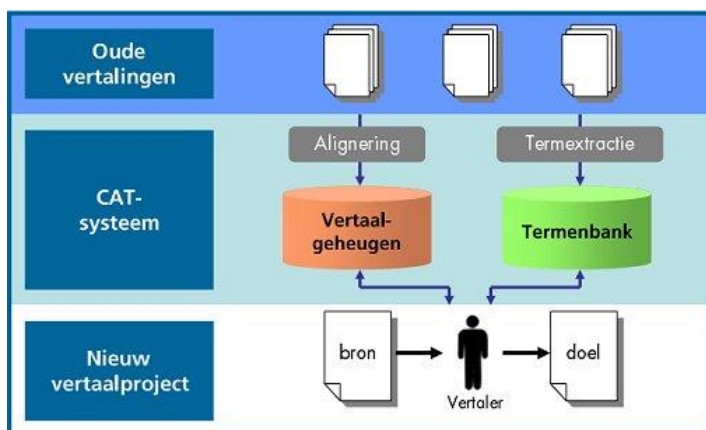
⁴⁰Arnold et al. 1994. *Machine Translation: An Introductory Guide*. 28 maart 2012 <<http://www.ling.ohio-state.edu/~dickinso/384/wi04/slides/mt-slides-4up.pdf>>.

3 Computerondersteund vertalen

Computerondersteund vertalen komt van het Engelse *Computer-assisted translation* of *Computer-Aided Translation*, afgekort tot CAT. Het gaat om computerprogramma's die het vertaalproces ondersteunen en vergemakkelijken. Het is een vertaalproces waarbij een menselijke vertaler gebruik maakt van software om een grotere precisie en efficiëntie te bereiken.

Belangrijke begrippen zijn hierbij het *vertaalgeheugen* en de *termenbank* (*glossarium of terminologielijst*). Het vertaalgeheugen is een database waarin segmenten, d.w.z. meestal zinnen, worden opgeslagen met hun vertaling, die dan voor nieuwe vertaalprojecten gebruikt kunnen worden. Een belangrijk onderdeel van een professionele vertaling is het consistent toepassen van de juiste terminologie. Om de juiste terminologie te krijgen zijn vertaalgeheugens vaak niet genoeg, omdat ze op zinsniveau werken en woorden kunnen daar niet uit gehaald worden. Daarom zijn er terminologielijsten die vak- of klantspecifiek kunnen zijn. Bepaalde termen en ook productnamen worden samen met hun vertaling in een terminologielijst vastgelegd. Op die manier stelt het CAT-programma keer op keer voor wat de vertaling van een bepaalde term moet zijn. In het vertaalproces wordt de terminologielijst gekoppeld aan een vertaalgeheugen waardoor de vertaler tijdens het vertalen automatisch de juiste terminologie krijgt. Op deze wijze wordt de tekst consequent met de juiste termen vertaald wat de kwaliteit van de vertaling verhoogt.⁴¹

Afbeelding 3 Het proces van vertalen met een CAT tool⁴²



⁴¹“Een terminologielijst samenstellen“. November 2011. Tilburg translation. 18 februari 2012 <<http://www.tilburgtranslations.nl/tip-van-de-maand/>>.

⁴²Het schema is beschikbaar op de volgende pagina: <<http://users.skynet.be/xenotext/nl/computerondersteund-vertalen-cat.htm>>.

*CAT-tool includes translation tools, linguistic tools and localization tools such as translation memory systems, electronic dictionaries or glossaries. It means that translator can use a broad range of various tools during a translation to make his/her work more efficient. The computer becomes a 'workstation', in which the translators create their own working environment thanks to access to these tools or programs.*⁴³ Naast vertaalgeheugen en terminologielijsten bevatten programma's spellingcontrole en vaak ook een geïntegreerd programma voor machinevertalen of lokalisatie software (zie hoofdstuk 5).

3.1 Vertaalgeheugen

Vertaalgeheugenprogramma's slaan vertaalde zinnen of tekstsegmenten op in een database zodat ze later eenvoudig opnieuw gebruikt kunnen worden. Door het vertaalgeheugen kunnen vertalingen sneller, goedkoper en consistentier geproduceerd worden dat wil zeggen dat repetitieve teksten steeds op dezelfde manier worden vertaald. Dit geeft het gebruik van consistente terminologie binnen een project. Dit speelt vooral bij updates van documentatie.⁴⁴

3.1.1 Principe van vertaalgeheugens

De tekst wordt verdeeld in segmenten, gewoonlijk is één segment een complete zin. Zodra een segment vertaald is, wordt het opgeslagen in het vertaalgeheugen. Als een identiek of vergelijkbaar segment opnieuw voorkomt in dezelfde of in een anderetekst, stelt het vertaalgeheugen automatisch de opgeslagen vertaling van dit segment voor. De vertaler kan dan de voorgestelde vertaling accepteren, verwerpen of wijzigen.

Het gebruik van een vertaalgeheugen werkt niet voor literaire teksten, wegens de creatieve taal die daar noodzakelijk is. Het vertaalgeheugen is geschikt voor het vertalen van repetitieve teksten met vaste terminologie bijvoorbeeld technische of juridische teksten.⁴⁵

⁴³Chiew K. Quah. *Translation and Technology* (Basingstoke: Palgrave Macmillan, 2006), 174.

⁴⁴Peter van Bart, M. Steehouder. *Basisboek: Technische Communicatie* (Assen: Koninklijke V.G., 2008) 292.

⁴⁵Vertaalgeheugens". 2011. Industrial services. 5 februari 2012 <http://www.be.sgs.com/nl_be/translation-memories.htm?selen=1&serviceid=10143648&lobid=5550>.

3.1.2 Exact match vs. Fuzzy match

Het vertaalgeheugen kan worden gezien als een database van teksten met vertalingen. Het programma vergelijkt segmenten van de nieuwe brontekst met de inhoud van het vertaalgeheugen. Repetitieve zinnen of zinsdelen worden automatisch vertaald en nieuwe vertalingen worden aan het systeem toegevoegd. Bij het vertalen van een nieuwe tekst zoekt de machine de *matches*. Een vertaler bepaalt zelf met welke graad van *fuzziness* kan worden gezocht.⁴⁶

- **Exact match**

Een opgeslagen bronsegment in een vertaalgeheugen is 100% identiek met een nieuw bronsegment in de te vertalen tekst.

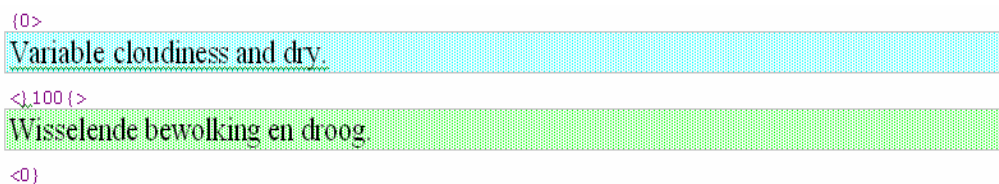
- **Fuzzy match**

Een opgeslagen bronsegment in een vertaalgeheugen lijkt op een nieuw bronsegment in de te vertalen tekst. De aanbevolen minimale *match* is 70%. Als het percentage lager is, komen de vertalingen vaak te weinig overeen omdat die dangebaseerd is op hoogfrequente grammaticale woorden zoals *en*, *de* of *van*.

De afbeelding nr 4 is een voorbeeld van een *exact match* (in het programma Wordfast) wat betekent dat de zinnen geheel overeenkomen. Het te vertalen document wordt gesegmenteerd en vergeleken met eerder vertaalde segmenten die in een vertaalgeheugen opgeslagen zijn. Indien Wordfast een exact match vindt, verschijnt de eerdere vertaling als voorstel in het document tegen een groen gekleurde achtergrond.⁴⁷

Afbeelding 4 *Exact match*

```
{0>  
Variable cloudiness and dry.  
<100 {>  
Wisselende bewolking en droog.  
<0}
```



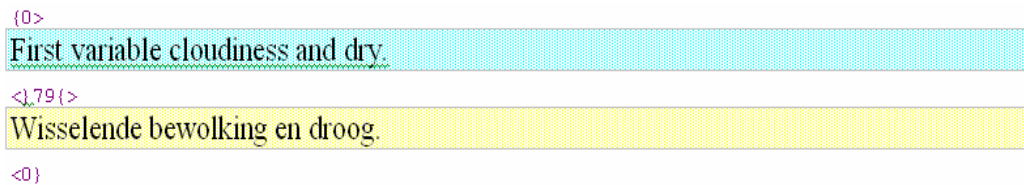
De afbeelding nr 5 is een voorbeeld van een Wordfast *fuzzy match*. Het matchpercentage is in dit geval 79%, dus er is slechts een klein verschil tussen beide zinnen (in dit geval ontbreekt een vertaling van het woord *first*). De fuzzy match verschijnt als voorstel tegen een geel gekleurde achtergrond. De vertaler kan de voorgestelde vertaling nu direct accepteren of het doelsegment eerst nog redigeren

⁴⁶Gerrit Sanders. 2009. "Computeronderstend vertalen". <<http://users.skynet.be/xenotext/nl/computerondersteund-vertalen-cat.htm>>.

⁴⁷De kleuren kunnen door een vertaler aangepast worden.

voor hij doorgaat naar het volgende segment. Elke nieuwe vertaling wordt weer opgeslagen in het vertaalgeheugen.⁴⁸

Afbeelding 5 Fuzzy match

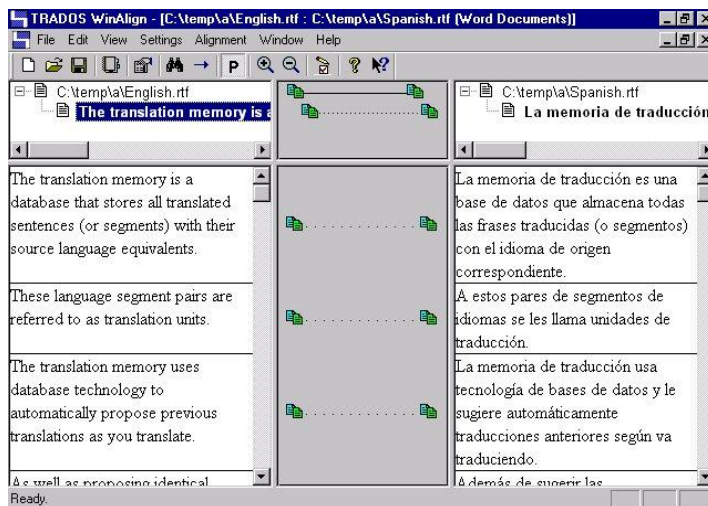


3.1.3 Alignering

Alignering van het Engelse *alignment*, is een proces waarbij gespecialiseerde software een brontekst met zijn vertaling vergelijkt en equivalente segmenten aan elkaar koppelt. Gekoppelde zinnen kunnen daarna naar het vertaalgeheugen worden geëxporteerd, een voorbeeld van zulke software is WinAlign van DLS Trados (zie onderstaande afbeelding nr 6).

WinAlign stelt zelf combinaties van bron- en doelsegmenten voor, door middel van stippelijntjes. Als een vertaler het niet eens is met een suggestie, kunnen de segmenten manueel worden samengevoegd of gesplitst. Als alle bron- en doelzinnen op de juiste manier aan elkaar gekoppeld zijn, kan het projectbestand naar het vertaalgeheugen geëxporteerd worden.⁴⁹

Afbeelding 6 WinAlign



⁴⁸Vertalen met behulp van CAT-software“. Eloquo. 5 april 2012
<<http://www.eloquo.nl/computer-aided-translation.html>>.

⁴⁹“WinAlign“. 8 september 2007. Taggers help. 11 februari 2012
<<http://taggershelp.blogspot.com/2007/09/winalign.html>>.

3.1.4 Manieren om vertaalgeheugen te verschaffen

Een nieuw vertaalprogramma is leeg, het bevat geen vertalingen. Er zijn verschillende manieren om een database te vullen:

- Zelf toevoegen aan de database van het programma:
 - Tijdens het vertalen het programma gebruiken: het bron- en doeltaal-segment worden aan elkaar gekoppeld en automatisch aan het vertaalgeheugen toegevoegd.
- Het importeren van eerder gemaakte vertalingen:
 - Met behulp van een align-programma: parallelle bestanden in bron- en doeltaal in segmenten gedeeld en aan elkaar gekoppeld. Later kunnen ze geïmporteerd worden in het vertaalgeheugen.
 - Export van TM van derden (meestal van klanten of vertaalbureau) importeren in het vertaalgeheugen.⁵⁰

3.2 *Voordelen en nadelen van CAT programma's*

In dit hoofdstuk worden enkele voordelen en nadelen van vertaalprogramma's besproken. Het is echter niet mogelijk om te generaliseren. De opinies van de vertalers verschillen van elkaar, niet alle vertalers gebruiken CAT tools, sommigen vinden ze niet nodig. De opinie hangt ook van het standpunt af: wat een voordeel voor de een kan betekenen, kan een nadeel voor anderen zijn. Een goed voorbeeld is bijvoorbeeld de prijs: vertaalbureaus bieden voor de repetitieve woorden een lager tarief aan en klanten verwachten dat ook. Voor vertalers betekent dat ze minder geld voor een vertaling krijgen en meer moeten vertalen om hetzelfde bedrag te verdienen.

3.2.1 Voordelen

- *Volledigheid*: de meeste systemen met vertaalgeheugens accepteren geen onvertaalde segmenten dat wil zeggen dat alles moet vertaald worden. Bij het vertalen zonder CAT tools gebeurt vaak dat een vertaler een zin of een alinea per ongeluk overslaat of verwijdert. Dat kan met vertaalgeheugens niet gebeuren omdat het programma herkent dat de vertaling niet volledig is.

⁵⁰“Vertaalgeheugensprogramma's”. Vertalers Platform. 18 februari 2012 <<http://www.vertalersplatform.nl/iv-vertaalgeheugenprogs.php>>.

- *Tijdbesparing*: hoe meer zinnen eerder vertaald zijn en uit de database kunnen worden overgenomen, hoe sneller een vertaling geleverd kan worden. Dat betekent dat voor de vertaling benodigde termijnen aanzienlijk worden verkort. Aan de andere kant is het een nadeel voor vertalers die dagelijks meer pagina's moeten vertalen. Concurrentievermogens worden noodzakelijk, vertalers die zonder CAT tools werken, kunnen niet aan de vereiste snelheid voldoen.
- *Lagere vertaalkosten*: doordat bestaand materiaal opnieuw wordt gebruikt, zijn de vertaalkosten lager – eerder vertaalde zinnen worden tegen een lager tarief aangeboden. Aan de andere hand is het weer een nadeel voor vertalers omdat ze voor zich herhalende segmenten lagere tarief krijgen en om dezelfde bedrag te verdienen moeten ze dagelijks meer pagina's vertalen.
- *Uniformiteit van de tekst*: zinnen die eerder zijn vertaald, worden bij updates en bij nieuwe versie van het document op dezelfde wijze vertaald met dezelfde terminologie als ze uit de database worden gehaald. Dit is handig als bijvoorbeeld meerdere vertalers aan hetzelfde project werken of als een handleiding is aangepast voor een nieuwere versie van het product.

Uit de bovenstaande punten blijkt dat deze voordelen relatief zijn. Wat een voordeel voor een klant kanleveren, namelijk lagere prijs, vertaling klaar binnen kortere termijn, kan een nadeel voor een vertaler betekenen, namelijk een vertaler moet op hoogte zijn van de nieuwe technologie, deze technologieën aanschaffen en kunnen gebruiken. Verder moet hij/zij ook meer tekst voor een lager tarief vertalen.

*Verdere pluspunten:*⁵¹

- Het vertaalgeheugen kan niet alleen worden gebruikt met Word-documenten, maar ook met andere bestandsformaten zoals HTML, PowerPoint, WordPerfect, FrameMaker, enz.
- Voor elke klant kan een apart vertaalgeheugen worden gecreëerd, zodat steeds de specifieke bedrijfsterminologie van de klant wordt gehanteerd en gehandhaafd.

⁵¹“Vertaalgeheugens“. 2011. Industrial services. 5 februari 2012 <http://www.be.sgs.com/nl_be/translation-memories.htm?selen=1&serviceid=10143648&lobid=5550>.

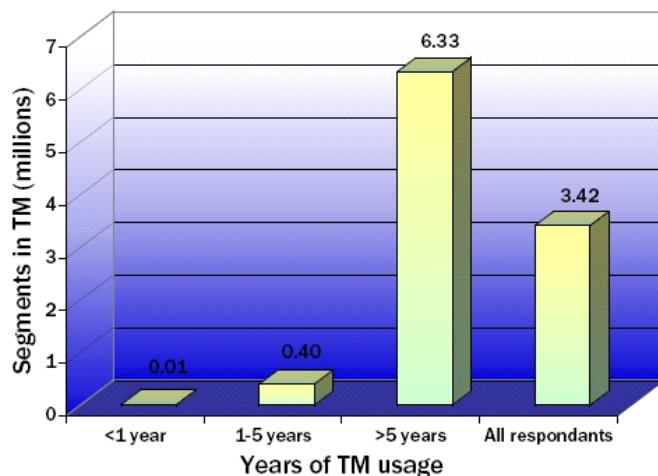
- Het vertaalgeheugen kan niet enkel worden gebruikt voor Europese talen, maar ook voor Fins, Russisch, Grieks, Oost-Europese talen en Aziatische talen (Chinees, Japans of Koreaans).

3.2.2 Nadelen

- *Niet voor alle teksten:* vertaalgeheugens zijn geschikt voor repetitieve teksten met een vaste terminologie. Het creatief schrijven is niet mogelijk dus geen literaire vertalingen.
- *Prijs:* de prijs van CAT-tools is redelijk hoog. Vaak moeten vertalers meer programma's aanschaffen omdat die door verschillende vertaalbureaus of klanten vereist worden.
- *Beperking:* vertalers worden beperkt in de vrijheid om hoe ze een tekst kunnen vertalen of wijzigen. Het document wordt zin voor zin vertaald, wat indruist tegen het primaire vertaalprincipe dat de tekst als een geheel moet worden vertaald niet als afzonderlijke zinnen. Het vertalen op zinsniveau maakt het soms moeilijker om het geheel in de gaten te houden. Het is natuurlijk heel belangrijk dat een vertaler de eindversie van de tekst uitprint en de doelttekst (vertaling) beoordeelt op logisch lopende zinnen, grammatica en lay-out.⁵²
- *Technischevermogens:* vertalers moeten op hoogte zijn van nieuwe technologieën en die kunnen gebruiken en benutten. Een nadeel is dat niet alle bestandsindelingen worden ondersteund bijvoorbeeld PDF of gedrukte tekst, het is noodzakelijk dan gebruik te maken van OCR (Optical Character Recognition) zie hoofdstuk 5.1.
- *Kwaliteit:* als vertaalgeheugens niet goed bijgehouden worden, kan het gebeuren dat foutieve vertalingen in de database opgeslagen worden die later opnieuw gebruikt worden.
- *Vertaalgeheugen moet opgebouwd worden:* het kan gebeuren dat een vertaler vertaalgeheugens van een klant of een vertaalbureau krijgt, maar als dat niet het geval is, moet hij/zij eigen vertaalgeheugens opbouwen. In het begin kan hij/zij er weinig gebruik van maken omdat de database weinig segmenten bevat. Vertaalgeheugens moeten dus opgebouwd worden. Dat wil zeggen dat

⁵²Ineke Kuiper. 2008. "Vertaalprogramma's". 18 februari 2012
<<http://translations.inekekuiper.nl/vertaalprogramma%27s.htm>>.

Figuur 1 *Vertaalgeheugen onderzoek*⁵³



hoe langer een vertaler vertaalgeheugen gebruikt en hoe meer teksten hij/zij vertaalt hoe meer segmenten zullen ze bevatten.

3.3 Terminologie

De terminologie is de studie van vaktaal en meer bepaald van de benamingen die binnen een vakgebied worden gegeven aan de voor het vakgebied relevante concepten. Die benamingen worden termen genoemd. Termen kunnen naar de vorm samenvallen met woorden of uitdrukkingen in de gewone omgangstaal maar binnen het vakgebied hebben zij een specifieke betekenis die niet altijd overeenkomt met de betekenis in de omgangstaal.⁵⁴ Bijvoorbeeld de juridische taal staat wat woordenschat betreft dicht bij de gewone omgangstaal. Om een paar voorbeelden te noemen: worden als ‘persoon’ of ‘schuld’ zijn vaktermen binnen de rechtswetenschap, en hebben binnen die wetenschap een betekenis die afwijkt van de betekenis van diezelfde woorden in de gewone omgangstaal.

3.3.1 Terminologielijst

Als er vaak dezelfde termen moeten worden vertaald, bijvoorbeeld bij een handleiding of bij technische teksten, is het belangrijk om bepaalde termen steeds op dezelfde manier te vertalen. Het is tijdrovend en vaak niet accuraat om dat handmatig of uit het hoofd te doen. Tegenwoordig bestaan er diverse software-

⁵³Vertaalgeheugen onderzoek gericht in 2002 door LISA – Localisation Industry Standards Association <<http://www.lisa.org/>>.

⁵⁴Gudrun Dawoens, *Taal aan den lijve: Het gebruik van corpora in taalkundig onderzoek en taalonderwijs* (Gent: Academia Press: 2008) 50.

programma's die de vertalers assisteren bij het consistent toepassen van de juiste terminologie. Dergelijke programma's maken gebruik van terminologielijsten. Er zijn enkele manieren om zo'n lijst samen te stellen. De meest eenvoudige vorm is een MS Excel-bestand van twee kolommen (zie afbeelding nr 7) of MS Word-bestand van twee kolommen die door Tab-toets gesplitst worden (zie afbeelding nr 8). Links de brontaal, rechts de juiste vertaling. Er zijn ook meer uitgebreide mogelijkheden, bijvoorbeeld een lijst waarbij een afbeelding aan een bepaalde term kan worden gekoppeld of waarbij een term uitgebreid kan worden beschreven en bijvoorbeeld ook informatie over het vakgebied kan bevatten.⁵⁵

Afbeelding 7 MS Excel terminologielijst

1	Term (GER)	Gender	Term (ENG)	Client
2	Abflüsse	m.	runoff	City of London
3	Abflusskomponenten	f.	runoff components	City of London
4	Absetzbecken	n.	tailing dam	City of London
5	Abwasser	n.	effluent, waste water	City of London
6	Aquifer	m.	aquifer	City of London
7	Aufbereitung	f.	water treatment	City of London
8	Ausflockung	f.	flocculation	City of London
9	Auslaugung	f.	exhaustion	City of London
10	Biotop	n.	biotope	City of London
11	Brauchwasser	n.	water for industrial use	City of London

Afbeelding 8 MS Word terminologielijst

Term (GER)	Gender	Term (ENG)	Client
Abflüsse	m.	runoff	City of London
Abflusskomponenten	f.	runoff components	City of London
Absetzbecken	n.	tailing dam	City of London
Abwasser	n.	effluent, waste water	City of London
Aquifer	m.	aquifer	City of London
Aufbereitung	f.	water treatment	City of London
Ausflockung	f.	flocculation	City of London
Auslaugung	f.	exhaustion	City of London
Biotop	n.	biotope	City of London

⁵⁵“Een terminologielijst samenstellen“. November 2011. Tilburg translation. 18 februari 2012 <<http://www.tilburgtranslations.nl/tip-van-de-maand/>>.

4 Vergelijking van de vertaalprogramma's

Er zijn veel vertaalprogramma's beschikbaar op de markt. Tot de bekendste en de meest gebruikte behoren Wordfast, SDL Trados, Déjà vu, MemoQ, daarnaast bijvoorbeeld Star Transit of Across, en verder ook Omega T en Google Toolkit die op het internet gratis beschikbaar zijn.

In dit hoofdstuk zullen de programma's Wordfast, SDL Trados, Déjà vu, MemoQ en Google Toolkit worden voorgesteld en vergeleken op grond van hun prijs, testperiode, ondersteunde talen en bestandsformaten enz. Google Toolkit wordt apart behandeld, omdat het een gratis beschikbaar programma is dat vooral door beginnende en niet professionele vertalers wordt gebruikt. Om de informatie over de programma's te verzamelen werd gebruik gemaakt van de officiële webpagina's van bepaalde vertaalprogramma's, verder ook van hun technische handleidingen en van andere commerciële bronnen en artikelen.

4.1 *Word-based vs. standalone (gebruiker interface)*

Als het programma *word-based* is, betekent het dat het binnen Microsoft Word en andere Microsoft Office documenten zoals Microsoft Excel of Microsoft PowerPoint werd geïntegreerd. Het voordeel daarvan is een intuïtief gebruik omdat Microsoft Word gebruikervriendelijk, bekend en overzichtelijk is. Hierdoor houdt de vertaler alle functies van Word tot zijn beschikking (bijvoorbeeld macro's en spellingscontrole) terwijl er een groot aantal bijzonder handige functies wordt toegevoegd, zoals de mogelijkheid om termen op te zoeken in woordenboeken, woordenlijsten, andere Office-toepassingen en op het internet. Het grootste nadeel is echter dat er problemen kunnen zijn met compatibiliteit met het oorspronkelijke programma (meestal Microsoft Word) bijvoorbeeld als er identieke functietoetsen zijn.

Een *Stand-alone* programma gebruikt een eigen interface. Een voordeel is de grotere stabiliteit. Het is een zelfstandig programma dat om deze reden geen problemen met compatibiliteit heeft. Sommige gebruikers kunnen echter het hanteren van een nieuwe interface lastig vinden en geven daarom liever de voorkeur aan geïntegreerde vertaalprogramma's. *Choosing a stand-alone versus an integrated interface is about making trade-offs. With the stand-alone approach you get greater stability and customization at the expense of feature richness,*

while with the integrated approach you get a rich feature set, document compatibility, and user familiarity at the expense of greater fragility.⁵⁶

4.2 Aangeboden producten

Tabel 4 Aangeboden producten SDL Trados

Translation Memory	Terminology Management	Software Lokalisatie	Machine Vertalen
SDL Trados Studio 2011	SDL MultiTerm Extract 2011	SDL Passolo 2011	SDL Enterprise Translation Server
SDL Trados Studio 2011 Starter Edition	SDL MultiTerm Desktop 2011		Language Weaver
SDL Trados Studio 2011 Freelance			Google Translate
SDL Trados Studio 2011 Professional			

SDL Trados biedt aantal verschillende producten aan (zie tabel 4). Het bekendste programma is SDL Trados Studio 2011. Een vertaler kan uit vier verschillende types van SDL Trados Studio kiezen aan de hand van de eigen behoeftes en voorkeuren. SDL Trados biedt ook twee terminologie-tools aan, namelijk SDL MultiTermExtract 2011 *data automatically locate and extract potential monolingual or bilingual terms from existing documentation or translation memories to build termbases and glossaries quickly*⁵⁷ en SDL MultiTermDesktop 2011 dat de capaciteit heeft om een onbeperkt aantal termen in een onbeperkt aantal talen te bewaren. SDL MultiTermDesktop 2011 is compatibel met andere bestandsformaten (bijvoorbeeld Microsoft Excel) en terminologieformaten (Term-Base eXchange of OLIF) wat het delen van terminologie met anderen gemakkelijker maakt.⁵⁸ SDL Passolo 2011 is een localisation tool die stand-alone kan zijn of geïntegreerd in andere SDL producten zoals SDL Trados en SDL MultiTerm.⁵⁹ SDL Trados Studio ondersteunt verder ook machinevertalen dat geïntegreerd kan zijn in alle producten. *SDL Trados Studio supports 3 machine translation engines that are available over an internet connection – SDL*

⁵⁶“Stand-alone versus embedded CAT tools: trade-offs“ 6 november 2008. Felix blog. 19 maart 2012 <<http://felix-cat.com/blog/2008/11/06/stand-alone-versus-embedded-cat-tools-trade-offs/>>.

⁵⁷“SDL MultiTerm Extract“. SDL. 8 april 2012 <<http://www.translationzone.com/en/translator-products/sdlmultitermextract/>>.

⁵⁸“SDL MultiTerm Desktop“. SDL. 8 april 2012 <<http://www.translationzone.com/en/translator-products/sdlmultitermdesktop/>>.

⁵⁹“SDL Passolo 2011“. SDL. 8 april 2012 <<http://www.translationzone.com/en/translator-products/sdl-passolo/>>.

Enterprise Translation Server,⁶⁰ Language Weaver,⁶¹ and Google Translate. By default, these provide a free of charge.⁶²

Wordfast biedt drie verschillende vertaalprogramma's aan (zie tabel 5):

Tabel 5 *Aangeboden producten Wordfast*

Translation Memory Tool	Applications
Wordfast Pro (WFP)	Wordfast Server
Wordfast Classic (WFC)	VLTM
Wordfast Anywhere (WFA)	

*Wordfast Pro offers all of the benefits you've come to expect from Wordfast products, including speed, ease-of-use, and value for your technology investment. Users can now take advantage of these benefits in a stand-alone, platform independent, java-based editing environment.⁶³ Ze bieden ook de mogelijkheid aan om machine vertalen te integreren. Wordfast Classic wordt geïntegreerd in Microsoft Word wat het gemakkelijk maakt om te leren en gebruiken. *Because Wordfast Classic operates entirely inside of Microsoft Word, there is no need to install or run any additional applications or extensions. The Microsoft Word interface decreases the learning curve and enables most users to be up and running in just a few hours. Users also benefit from the many features and shortcuts already available in Microsoft Word.*⁶⁴ Wordfast biedt verder ook zijn mega-vertaalgeheugen aan onder de naam *Wordfast Anywhere*. Het is gratis en online beschikbaar. Om het vertaalgeheugen te mogen raadplegen, moet een vertaler wel een (gratis) account aanmaken. *Wordfast Anywhere offers unlimited private Translation Memory + private Terminology + optional public TM + optional Machine Translation. The confidentiality of your data (documents, TM, terminology) is total and uncompromising: it's not shared (unless you decide**

⁶⁰Built on reliable, statistically based translation technology to support unique needs and language requirements of government organizations.

⁶¹Built on reliable, statistically based technology, the company's products have been deployed worldwid. SDL and Language Weaver make it easier for large organizations and governments to communicate effectively with customers, regardless of language.

⁶²“Automated Translation. SDL. 8 april 2012<<http://www.translationzone.com/en/translator-solutions/automated-translation/>>.

⁶³“Wordfast Pro – The next step in translation Memory technology. Wordfast. 25 maart 2012“<http://www.wordfast.com/docs/WF_ProProductbrief.pdf>.

⁶⁴“Wordfast Classic“. 2012. Wordfast. 27 maart 2012 <http://www.wordfast.com/products_wordfast.html>.

to share data with selected colleagues.⁶⁵ VTML staat voor Very Large Translation Memory. Het is a *server-based TM environment, accessible over the internet. The VLTM may only be accessed from Wordfast's desktop TM application, and is a feature not available with other CAT tools.*⁶⁶ Wordfast server is geschikt voor een collaboratieve werk tussen vertalers. Het programma kan tot één of meer vertaalgeheugens tegelijkertijd verbonden worden.

Tabel 6 Aangeboden producten Déjà Vu

Translation Memory Tools	Applications
Déjà Vu X2 Standard	Déjà Vu X2 Editor
Déjà Vu X2 Professional	Déjà Vu X2 TeaM Server
Déjà Vu X2 Workgroup	

Déjà Vu X2 biedt drie verschillende vertaalgeheugens aan (zie tabel 6). Déjà Vu X2 Standard is geschikt voor kleinere vertaalprojecten. Déjà Vu X2 Professional is het meest populair van de Déjà Vu X2 producten omdat het een brede reeks van functies en componenten verleent. Déjà Vu X2 Workgroup is geschikt voor vertaalbureaus en bedrijven. Verder Déjà Vu X2 TeaM Server *by integrating with Déjà Vu X2 Workgroup, it allows translators who work on extensive, multinational and multisite translation projects to efficiently and seamlessly share their translations in real-time, ensuring superior quality and consistency.*⁶⁷ Déjà Vu X2 Editor biedt aan vertalers die Déjà Vu X2 niet bezitten een mogelijkheid om gratis aan een Déjà Vu X2 Workgroup deel te nemen.

⁶⁵“Wordfast Anywhere“. 2012. Wordfast. 27 maart 2012 <http://www.wordfast.net/?which_page=anywhere>.

⁶⁶“Very Large Translation Memory Project“. 2012. Wordfast. 27 maart 2012<http://www.wordfast.com/products_vltm.html>.

⁶⁷“Déjà Vu TeaM Server“. 2012. Atril. 27 maart 2012<<http://www.atril.com/en/software/team-server>>.

Tabel 7 *Aangeboden producten MemoQ*

Translation Memory Tool	Applications
MemoQ Translator Pro	MemoQ server
	qTerm
	TM Repository

MemoQ biedt één vertaalgeheugen programma aan (zie tabel 7). De MemoQ server is geschikt voor coöperatieve projecten als twee of meer vertalers aan hetzelfde project werken. *A memoQ server allows translators to work together, to share every single sentence translated, every single term suggested in real time.*⁶⁸ qTerm is geschikt voor bedrijven en organisaties. Het helpt technische communicatie te vereenvoudigen en misverstanden te vermijden, verder ook kosten te verlagen. *qTerm allows companies and institutions to provide a single point of access to any terminology in the organization, and allows their translation providers to have immediate access to the most up-to-date terminology.*⁶⁹ TM Repository is een vertaalgeheugen management systeem dat helpt bedrijven, vertaalbureaus, instellingen en anderen om volledig controle te hebben over de vertalingen (al vertaalde teksten maar ook nog te vertalen teksten).

4.3 Prijs

In benedenstaande tabel nr 8 zijn de prijzen van besproken CAT tools vergeleken. Er zijn grote verschillen tussen de prijzen van de afzonderlijke producten. Trados is het duurste programma en Wordfast is aan de andere kant het goedkoopste. Programma's kunnen op officiële pagina's aangeschaft worden. Verder worden ze verkocht ook op de pagina www.ProZ.com. Het is een toonaangevend webportaal voor vertalers. Leden van ProZ kunnen speciale aanbiedingen op TGB⁷⁰ benutten. *TGB is a tool offered to give an opportunity for translators buying CAT tools and other software together to get a lower price and realize significant savings through group buying.*⁷¹ De besparing kan zelfs meer dan 50% van de prijs zijn.

⁶⁸“MemoQ Server“. 2012. Kilgray Translation Technologies. 28 maart 2012 <<http://kilgray.com/products/memoq-server>>.

⁶⁹“qTerm“. 2012. Kilgray Translation Technologies. 28 maart 2012 <<http://kilgray.com/products/qterm>>.

⁷⁰TGB = Translator Group Buying.

⁷¹“Translator Group buying“. 2012. ProZ. 9 april 2012 <http://www.proz.com/?sp=store&sp_mode=overview>.

Tabel 8 Prijs van de programma's

Translation Memory Tool	Prijs (officiële webpagina)	TGB op ProZ (25.3.2012)
SDL Trados Studio 2011 Starter Edition	€99.00 (introductory price)	
SDL Trados Studio 2011 Freelance	845 €	549 € (+ a year membership ⁷²)
SDL Trados Studio 2011 Professional	2 895€	
SDL MultiTerm Extract 2011	400€	
SDL Passolo 2011	1940€	
Wordfast Pro	350 €	325 € (+ a year membership + 3 hour online training ⁷³)
Wordfast Classic	350 €	430 € (+ a year membership + 3 hour online training)
Wordfast Studio (Classic + Pro)	500 €	
Déjà Vu X2 Standard	290 €	
Déjà Vu X2 Professional	690 €	345 € (membership not included)
Déjà Vu X2 Workgroup	1490 €	
MemoQ Translator Pro	620 €	372 € (+ a year membership)

4.4 *Systeembenodigheden*

Uit de benedenstaande tabel nr 9 blijkt dat alle CAT programma's gelijke besturingssystemen (operating systems) hebben die aanbevolen zijn voor programma'som optimaal te kunnen werken. Alle programma's zijn compatibel met Microsoft Windows XP en Vista. Ook oudere versies van Microsoft Windows 2003, 2008 (behalve Wordfast Pro) zijnmogelijk, verder ook Microsoft Windows 2000 en 1998 zijnondersteundmaar slechts door Wordfast Classic en MemoQ. Linux en Mac zijn compatibel alleen met Wordfast Classic en Wordfast Pro.

⁷²One year membership op ProZ.com 114 € (As a ProZ.com member, you will enjoy complete and immediate access to all site functions and data. This includes full access to jobs, KudoZ and the Blue Board database of outsourcers with member entries.)

⁷³Wordfast 3 uur online training door ProZ.com 76 €.

Tabel 9 *Systeembenodigheden*

CAT programma	Ondersteunde platformen
Déjà Vu X2	Microsoft Windows XP, Vista, 7 Windows Server 2003, 2003 R2, 2008, 2008 R2
MemoQ	Microsoft Windows 7, 2003, Me, 98, 2000, Vista, XP, Windows Server 2008
SDL Trados 2011	Miscrosoft Windows XP, Vista, 7 Windows Server 2003, 2008 (not supported – Windows 2000, NT, 98, XP pre-SP3, 2003 pre-SP2)
Wordfast Classic	Microsoft Windows 95, 98, ME, NT 4, 2000, XP, Vista, 7 Mac OS 7, 8, 9, OSX 10.3.4 en hoger Linux (Mandrake, XFCE, VMWare & Office 2007), Linux (Suse, KDE2 + Crossover & Word 2000)
Wordfast Pro	Windows XP, Vista, 7 Mac OSX 10.4 en hoger Linux

4.5 Testperiode (demoversie)

Vertalers hebben een mogelijkheid om met een programma te werken voor ze hem gaan aanschaffen dat wil zeggen dat ze een demoversie kunnen downloaden en proberen welk programma ze het beste vinden (zie tabel nr 10):

Tabel 10 *Testperiode*

Translation Memory System	Demo Version (30 dagen)
Déjà Vu	Déjà Vu X2
MemoQ	MemoQ server, qTerm, TM Repository
SDL Trados	SDL Trados Studio 2011 Professional, SDL Passolo 2011
Wordfast	Worfast Classic, Worfast Pro

Na de 30 dagen durende testperiode is het noodzakelijk om het programma te kopen om alle aangeboden functies verder te kunnen gebruiken.

Voor beginnende vertalers kunnen vooral Wordfast Classic en Wordfast Pro interessant zijn. Met een ongeregistreerde versie van Wordfast is het mogelijk om met vertaalgeheugens van maximaal 110 kB en/of 500 vertaaleenheden gratis te werken.

4.6 Ondersteunde talen

Wat de ondersteunde talen van SDL Trados, Wordfast, Déjà vu en MemoQ betreft, is het niet nodig om de vergelijking te maken omdat alle programma's talen ondersteunen die ook door Microsoft Windows worden ondersteund, namelijk:

- Het Latijnse alfabet – een alfabet dat voor vrijwel alle westerse spellingen wordt gebruikt zoals onder andere Engels, Duits, Frans, Nederlands of Tsjechisch
- Het cyrillische alfabet – een alfabet dat in zes Slavische talen gebruikt wordt, namelijk Russisch, Oekraïens, Servisch, Macedonisch, Bulgaars en Wit-Russisch
- Aziatische talen zoals Japanees, Chinees, Thai of Koreaans
- Rechts naar links geschreven talen zoals Arabisch, Hebreeuws
- Een aantal kleinere talen

Het betekent dat alle CAT programma's dankzij Unicode⁷⁴ van en naar alle talen kunnen vertalen. Er zijn ook voor gecompliceerde talen geen limieten. Dat zijn bijvoorbeeld van rechts naar links geschreven talen of talen met DBCS-tekensets (double byte character set) zoals onder meer Japans, vereenvoudigd Chinees, traditioneel Chinees en Koreaans, dat wil zeggen talen met hoog aantal speciale tekens en symbolen (één byte-tekenset heeft maximaal 256 tekens terwijl een twee byte-tekenset tot 65 536 tekens kan bevatten).⁷⁵

4.7 Ondersteunde bestandsformaten

Alle programma's ondersteunen MS Word, MS Excel, MS Powerpoint en HTML. Wordfast Pro en Classic zijn het meest gelimiteerd. Aan de andere kant Déjà Vu en SDL Trados ondersteunen de meest voorkomende bestandsformaten dat wil zeggen rond de zeventig formaten.

⁷⁴Unicode is een internationale standaard voor de codering van binaire codes naar grafische tekens en symbolen. De standaard voorziet alle tekens en symbolen van alle geschreven talen.

⁷⁵“Character sets”. 2012. PHP Community. 15 maart 2012
<<http://www.pfz.nl/wiki/charactersets/>>.

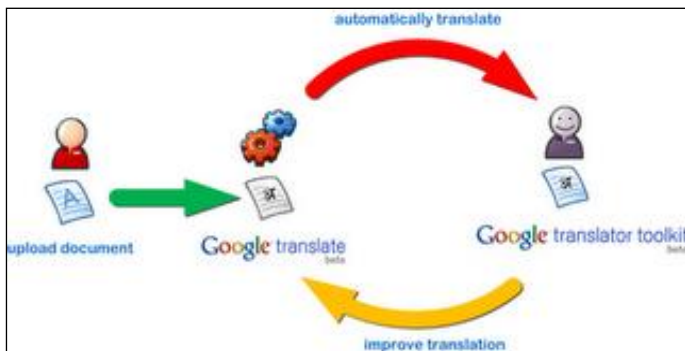
Tabel 11 *Ondersteunde bestandsformaten*

Translation Memory System	Ondersteunde bestandsformaten
Déjà Vu	MS Word, Rich Text Format, Windows Help, MS Excel, MS PowerPoint, MS Access, OpenOffice/StarOffice, Adobe FrameMaker MIF, Adobe InDesign, Adobe PageMaker tagged text, QuarkXPress XTG, Interleaf ASCII, HTML, ASP/ASP.NET, PHP, JavaScript, VBScript, HTML Help, SGML, XML, RC, C/C++/Java, Java, IBM TM niet-vertaalde segmenten, Trados WorkBench documenten, TradosTag TTX, Trados TagEditor BIF, Trados TagEditor TMX, GNU PO, POT files, en plain text files.
MemoQ	Adobe Framemaker, Adobe InDesign, Adobe InDesign, AuthorIT, DITA, FreeMind, HTML, Java properties, MS Word, MS PowerPoint, MS Visio, MS Help, OpenDocument text dokument, PDF, Plain text, Rich Text Format, Scalable Vector Graphics, TradosTag, TRADOS bilingual RTF, Typo3, ResX, XML, XLIFF
SDL Trados	meer dan 70 bestandsformaten, ook de nieuwste: Microsoft Office 2010, PDF, Adobe FrameMaker 9.0 and 10.0, Adobe InCopy CS5, Adobe InDesign CS5, OpenDocument en Google Docs
Wordfast Classic	MS Word, MS Excel, MS Powerpoint, HTML
Wordfast Pro	MS Word, MS Excel, MS Powerpoint, HTML, XML, ASP/JSP/Java, INX (InDesign), MIF (FrameMaker), TTX, PDF

4.8 Google Toolkit

Google Toolkit wordt apart behandeld omdat het een commercieel, gratis beschikbaar programma is dat vooral door niet-professionele vertalers gebruikt wordt. Het is een project van Google dat in juni 2009 werd geïntroduceerd. Gebruikers kunnen een gratis account aanmaken. Vervolgens kunnen ze projecten starten waarbij meerdere vertalers samenwerken door hetzelfde vertaalgeheugen te delen. Het systeem ondersteunt verschillende bestandsformaten en komt met een automatische vertaalmodule. In de benedenstaande afbeelding nr 9 is het proces van vertalen te zien. Een geüpload document wordt automatisch vertaald door Goole Translate, een gebruiker kan de automatisch vertaalde tekst vervolgens aanpassen en verbeteren.

Afbeelding 9 Google Toolkit – het proces van vertalen⁷⁶

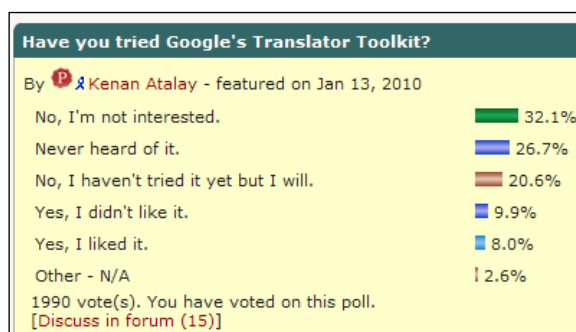


Programmavereisten:

- Één brondocument kan maximaal 1MB hebben en er kan maximaal 1GB vertalingen per jaar upgeload worden.
- Ondersteunde bestandsformaten:
AdWords Editor Archive (.aea), Android Resource (.xml), Application Resource Bundle (.arb), Chrome Extention (.json), GNU gettext (.po), HTML, Plain text (.txt), Microsoft Word (.doc), Rich-text (.rtf), Open Office (.odt) en Wikipedia.

Google Translator Toolkit ondersteunt Wikipedia. De Toolkit vereenvoudigt het vertalen van Engelstalige Wikipedia artikelen naar andere talen. Alle vertalingen van Wikipedia worden in een gemeenschappelijk vertaalgeheugen opgeslagen dat voor iedereen beschikbaar is. Wat kan de tijd besparen als iemand begint een al vertaald artikel te vertalen.⁷⁷

Figuur 2 Have you tried Google's Translator Toolkit? (ProZ)



⁷⁶De afbeelding beschikbaar op de volgende pagina: <<http://www.findmysoft.com/news/Google-Translator-Toolkit-Ads-Human-Touch-to-Google-Translate/>>.

⁷⁷Tony Ruscoe. Juni 2009. "Google translator Toolkit". 6 maart 2012 <<http://blogscoped.com/archive/2009-06-09-n19.html>>.

In de bovenstaande figuur nr 2 zijn de opinies van vertalers op ProZ over Google Translate Toolkit te zien. Ze tonen echt een kleine interesse in dit programma. Slechts 20,6% wil het proberen en alleen 8% vindt hem goed. Vertalers zijn sceptisch jegens Google en vertrouwen het niet.

Er zijn enkele nadelen die misschien niet direct zichtbaar zijn. Google Translator verzamelt data, hoe meer data hoe beter de vertalingen zullen worden. In het geval van Toolkit betekenen die data gezamenlijke vertaalgeheugens en glossaria. Wie Toolkit gebruikt “deelt“ eigen vertalingen met Google. In het geval dat 100,000 vertalers Toolkit zou gebruiken, krijgt Google 50 biljoen vertaalde woorden van goede kwaliteit binnen een jaar wat zal hem helpen om Google Translate te verbeteren.⁷⁸ *Google is a commercial enterprise and may decide that translation is a good source for additional revenue in the future. The meter could be switched on and translators will find themselves hijacked and punished for their naivety.*⁷⁹ Het kan de reden zijn waarom vertalers Google niet vertrouwen. Vertaalgeheugens en glossaria zijn door hard werken verkregen eigendom en er is geen reden waarom dat gratis zou worden verleend aan Google en breed publiek.

⁷⁸Jaap Van Der Meer. 17 juni 2009. „Google Translate Toolkit“. TAUS. 20 april 2012 <<http://www.translationautomation.com/technology/google-translation-toolkit.html>>.

⁷⁹Jaap Van Der Meer <<http://www.translationautomation.com/technology/google-translation-toolkit.html>>.

5 Verdere technologieën en begrippen noodzakelijk voor vertales van vandaag

Er zijn ook andere taaltechnologieën die nauw met computerondersteund vertalen verbonden zijn waarvan kennis noodzakelijk is om vandaag de dag vertaler te worden. Dat zijn bijvoorbeeld lokalisatie, machinevertalen, optische tekenherkenning of croudsourcen.

5.1 Optische Tekenherkenning

Optische tekenherkenning of in het Engels *Optical Character Recognition* afgekort tot OCR, is een technologie voor het maken van machine-bewerkbare tekst uit de afbeelding in bestanden van gescande documenten. Wanneer een document wordt gescand in een computer, wordt dit opgeslagen als een beeldbestand zoals PDF, TIFF of JPG. Zonder gebruik van OCR-software, is dat beeld alleen een beeld op de computer. Zo'n tekst heeft verschillende nadelen: de gebruiker kan geen woorden zoeken, noch ook de tekst bewerken. Dit is een technologie noodzakelijk voor vertalers die met CAT-tools en vertaalgeheugens werken. Als ze een PDF of gedrukt document krijgen, kunnen ze deze dankzij optische tekenherkenning in een bewerkbaar bestand omzetten. *The use of OCR software is growing amongst translators; one reason is no doubt the increasing use of translation memory applications, which require the source text to be in a machine-readable form.*⁸⁰

Het herkennen van handschriften is ook mogelijk. In het Engels wordt het tot ICR afgekort (Intelligent Character Recognition). Het herkennen handschrift-herkenning is ingewikkelder dan het herkennen van gedrukte tekst, vanwege de grote variatie in schrijfwijzen. Toch wordt deze techniek veel en succesvol ingezet voor het herkennen van handgeschreven teksten op vooral formulieren. Doordat de tekst daar op vaste plaatsen staat en ook vaak een specifieke betekenis heeft, bijvoorbeeld postcode, registratienummer, adres, getal, is er door de software-systemen toch een goede herkenbaarheid.⁸¹

⁸⁰Marc Prior. 2011. "Optical character recognition". 9 februari 2012 <<http://www.marcprior.de/linux/misc/ocr.html>>.

⁸¹"OCR". 2012. CVISION Technologies. 9 februari 2012 <<http://www.cvisiontech.com/reference/general-information/ocr.html?lang=dut>>.

5.2 Lokalisatie

De Engelse term *localisation* wordt soms tot het numeroniem L10N afgekort, omdat er tussen de L en de N nog tien letters zijn. De nood aan vertalingen is door de digitale revolutie toegenomen, voornamelijk in de lokalisatie-industrie, waar men uitgaat van de gedachte dat het product aan elke deelmarkt aangepast moet worden, en dat die plaatselijke markt dus ook zoveel mogelijk in zijn eigen taal bediend moet worden. Hier is het vertalen slechts een onderdeel van een veel omvattender proces van adaptatie van een product aan lokale markten.

De LISA (Localisation Industry Standards Association) definieert lokalisatie als volgt: *Het aanpassen en vertalen van een product of dienst in een andere taal met als doel het pakket linguïstisch, technisch en cultureel geschikt te maken voor een bepaalde lokale markt.* Bert Esselink beschrijft het proces van lokalisatie als volgt: *In a nutshell localization revolves around combining language and technology to produce a product that can cross cultural and language barriers. No more, no less.*⁸²

Het vertaalproces wordt hier gekenmerkt door veelvuldige updates van de brontekst waardoor gebruik van technologische hulpmiddelen (terminologische databanken, vertaalgeheugens en machinevertaling) nodig is.⁸³ Beschikbare lokalisatie programma's zijn bijvoorbeeld WebBudget XT, SDL Passolo, Alchemy Catalyst, SDL Trados of Multiterm. *These tools are used to translate software re-source files or even binary files and enable the localizer to not only translate but also resize and test the user interface.*⁸⁴

5.2.1 Waarom noodzakelijk

Het omzetten van een document naar een andere taal gaat vaak verder dan alleen vertalen. De documentatie moet bijvoorbeeld worden aangepast aan nationale standaarden en regelgeving, voorbeelden of illustraties moeten worden vervangen door een inhoud die beter past bij de lokale markt en cultuur waarvoor de vertaalde versie bedoeld is.

Lokalisatie is het aanpassen en vertalen van een product of tekst in een andere taal met als doel het eindproduct taalkundig, technisch en inhoudelijk geschikt te maken voor een bepaalde markt. Een gelokaliseerd eindproduct kan bijvoorbeeld

⁸²Bert Esselink. 2012. "Evolution of Localisation". 18 februari 2012 <<http://www.scribd.com/doc/12596514/The-Evolution-of-Localization>>.

⁸³VANDEWEGHE 55-56.

⁸⁴Bert Esselink. "Evolution of Localisation".

software zijn, de interface van een mobiele telefoon of televisie, een website of een handleiding.⁸⁵ *In addition to translation (and, therefore, grammar and spelling issues that vary from place to place where the same language is spoken), the localisation process might include adapting graphics; adopting local currencies; using proper forms for dates, addresses and phone numbers; the choices of colours; and many other details, including rethinking the physical structure of a product. All these changes aim to recognize local sensitivities, avoid conflict with local culture and habits, and enter the local market by merging into its needs and desires. For example, localisation aims to offer country-specific websites of the same company, or different editions of a book depending on the place it is published.*⁸⁶

Uit de bovenstaande definitie blijkt dat een vertaler niet alleen taal perfect moet beheersen maar hij/zij moet zich ook bewust zijn van culturele- en nationale verschillen, bijvoorbeeld in maten en getallen die van een brontaal tot een doeltaal juist omgezet moeten worden en voor de lezer begrijpelijk en natuurlijk gemaakt worden. De lezer mag niet merken dat hij/zij een vertaling leest. Er zijn enkele voorbeelden:

- **Decimale getallen**

In het Nederlands wordt de komma voor decimalen gezet, zoals in € 14,95 en \$ 140,00. Dat verschilt van Engelstalige landen, waar de punt het decimaalteken is: \$ 140.00. In het Nederlands wordt geschreven 3.525,75; in het Engels echter 3,525.75.

- **Maten**

Inhoudsmaat pint is 0,568261 l.

Een Engelse mijl is 1609,344 meter.

- **Datumformaat**

Datum formaat in de Verenigde staten heeft het formaat “Maand/Dag/Jaar” (mm/dd/yyyy). In het Nederlands en andere Europese landen is het echter “Dag/Maand/Jaar” (dd/mm/yyyy). Als de vertaler dit verschil niet merkt kan het in sommige gevallen tot misverstand leiden.

⁸⁵Peter van Bart, M. Steehouder, *Basisboek: Technische Communicatie* (Assen: Koninklijke Van Gorcum, 2008) 286.

⁸⁶“What is the difference between translation and localisation?” 2009. Translation services. 11 februari 2012 <<http://www.translation-service.com/faqs/localization/>>.

Tabel 12 *Verschillen in datumformaat*

NL	VS
20/06/2011	06/20/2011
05/06/2002	06/05/2002

Het is duidelijk dat de hedendaagse vertaler niet alleen een perfecte kennis van de bron- en doeltaal moet hebben maar ook over kennis over de cultuur en gewoonten van beide landen moet beschikken en op de hoogte moet zijn van nieuwe technologieën en hun ontwikkelingen.

- ***Verschillende lengte van de vertaling en de originele tekst***

In de meeste gevallen gebeurt het dat de vertalingen en de originele tekst verschillende lengte hebben (zie tabel nr 13). Het is belangrijk om in de gaten te houden dat de meeste talen langer zijn dan de originele Engelse tekst. Het is daarom noodzakelijk om in het originele bestand extra ruimte te voorzien. Er wordt aanbevolen om een ruimte van tenminste 30% extra vrij te houden. De opmaak zal naderhand slechts weinig aangepast hoeven te worden.⁸⁷ Het is noodzakelijk om te vermijden dat er teveel witte ruimte is of dat er te weinig plaats is en de vertaling afgekort moet worden. De lengte van de vertaling hangt van de stijl, soorttekst, onderwerp of zinnenconstructies af.⁸⁸

Tabel 13 *Verschillende lengte van de vertaling en de originele tekst*⁸⁹

Brontaal	Doeltaal	Lengte
Engels	Tsjechisch	+10%
Engels	Nederlands	+20%
Tsjechisch	Engels	van -5% tot -10%
Nederlands	Engels	-20%

⁸⁷ Aykin Nuray, *Usability and Internationalization of Information Technology* (NJ: Lawrence Erlbaum Associates, 2004) 25-26.

⁸⁸ Adam Wooten. 21 mei 2010. "Text Expansion & Contraction in Translation". The Official Localisation Blog of Globalization Group. 1 april 2012 <<http://www.globalization-group.com/edge/2010/05/text-expansion-contraction-in-translation/>>.

⁸⁹ Informatie beschikbaar op de volgende pagina:
<<http://www.globalization-group.com/edge/2010/05/text-expansion-contraction-in-translation/>>.

5.2.2 Globalisering / Lokalisatie / Vertalen

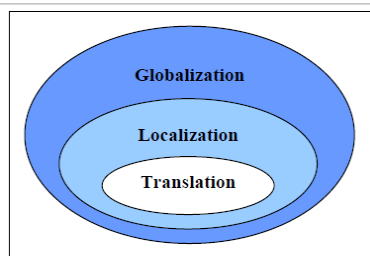
Vertaaltechnologie deskundige, Minako O'Hagan, gaf de volgende definitie:

Translation [...] can be placed as the core of both localization and globalization. [...] Localization [is now] treated as a new form of translation.

Hij is ook de auteur van de onderstaande diagram (afbeelding nr 10).⁹⁰

Globalisering,⁹¹ lokalisatie en vertalen kunnen we begrijpen als series van samen verbonden processen. Diagram beneden toont het wel aan – vertaling is de kern en de andere processen uitgaan van het.

Afbeelding 10 Globalisering, lokalisatie en vertalen⁹²



5.3 Machinevertaling

Machinevertaling wordt ook Automatische vertaling (AV) genoemd. De Engelse term is *Machine Translation* die vaak tot MT afgekort wordt. De ontwikkeling en geschiedenis van Machinevertaling is nauw verbonden met de ontwikkeling van de taaltechnologieën in het algemeen en om deze reden was dit al in het eerste hoofdstuk behandeld.

Het is belangrijk om te beseffen dat de machinevertalingen verre van perfect zijn en dat ze voor vertalers alleen op enkele gebieden van nut kunnen zijn. Sommigen vrezen wel eens de concurrentie van de machine voor de mens, maar wat er tot nu toe gebeurd is op het vlak van automatische vertaling, geeft aan dat

⁹⁰Minako O'Hagan. "Training for localisation". School of Applied Language and Intercultural Studies Dublin City University, Ireland. 5 februari 2012 <http://isg.urv.es/library/papers/OHagan_Training.pdf>.

⁹¹To *globalize* is to plan the design and development methods for a product in advance, keeping in mind a multicultural audience, in order to avoid increased costs and quality problems, save time, and smooth the localizing effort for each region or country.

⁹²Afbeelding beschikbaar op de volgende pagina: <http://isg.urv.es/library/papers/OHagan_Training.pdf>.

vertalers machines voor eigen profijt kunnen benutten en dat mens en machine complementair kunnen werken. *De belabberde resultaten van automatische vertaling op grond van een woordenboek gekoppeld aan een regelsysteem, leerden al snel dat machines niet in staat waren om te gaan met de benodigde achtergrondkennis om een tekst te begrijpen.*⁹³

Machinevertalingen kunnen van nut zijn bij repetitieve teksten met vaste terminologie, een goed voorbeeld daarvan is TAUM-Météo. Het is een automatisch vertaalsysteem voor weerberichten (EN>FR) ontwikkeld aan de universiteit van Montreal. Het succes van dit project ligt natuurlijk aan de tekstsoort waar stereotiepe wendingen steeds terugkeren, en de gebruikte woordenschat (hogedrukgebied, sneeuw, regen, enz.) die beperkt is. Machinevertaling kan ook gebruikt worden om een ‘ruwe’ vertaling van de tekst te krijgen. Dat wil zeggen om een idee te krijgen waar de tekst over gaat. Het is zeker niet geschikt om literaire teksten te vertalen omdat de computers talengevoel en wereldkennis missen. Computers kunnen ambiguïteit vaak niet oplossen en de morfologische en syntactische verschillen tussen talen kunnen ook fouten opleveren (deze problematiek werd al in het hoofdstuk twee besproken).

Ofschoon machinevertaling gebrekkig is, wordt het (vooral onder niet-vertalers) steeds meer populair. Het komt vooral door het internet. Al jaren zijn er zowel gratis als betaalde sites waar men een woord of zin in de ene taal invoert en binnen een seconde in de andere taal de vertaling ontvangt. Gratis beschikbare online programma’s zijn goed in het vertalen van woorden en zeer korte zinnen. Langere zinnen, uitdrukken, werkwoordvormen en grammatica worden vaak niet correct geïnterpreteerd en er wordt dus een tekst vol fouten opgeleverd. Machinevertalingen zijn handig om een goed beeld te krijgen waar een tekst over gaat, de strekking van het verhaal komt er meestal wel uit.⁹⁴

Er zijn twee methoden te onderscheiden: automatisch vertalen gebaseerd op regels (rule-based) of gebaseerd op statistiek (corpus-based). De bekendste programma’s die gratis op het internet aangeboden worden, zijn Google Translate, Babelfish en Bing die op verschillende principes werken. Google Translate is gebaseerd op statistiek en Babelfish en Bing op regels. Voorbeelden van betaalde programma’s zijn Babylon, Systran en Power Translator.

⁹³Vandeweghe 171-2.

⁹⁴“Google Translate: Gratis online vertalen met Google“. 1.4. 2010. PC and Internet: Tips and Tricks. 22 februari 2012 <<http://pc-en-internet.infonu.nl/tips-en-tricks/53497-google-translate-gratis-online-vertalen-met-google.html>>.

5.3.1 Rule-based MT

Veel van de moderne, commerciële systemen voor machinevertalingen die vandaag de dag gebruikt worden, zijn ontwikkeld op basis van regels. De computer leert de woordenboeken en grammatica's van verschillende talen. Deze kennis kan hij gebruiken om de tekst na een taalkundige analyse om te zetten in de doeltaal. Afhankelijk van de diepte van de taalkundige analyse zet hij de tekst woord voor woord, of per hele zin om. De gevormde zin is in ieder geval grammaticaal correct, maar wat de inhoud betreft zijn er vaak fouten. Dergelijke systemen zijn zeer bewerkelijk omdat een lexicon en een grammatica moeten worden gedefinieerd. De nadelen daarvan zijn daarom dat het veel tijd en geld kost. Op dit principe werkt bijvoorbeeld Babelfish of Bing.⁹⁵

5.3.1.1 Babelfish

Yahoo! Babel Fish is door AltaVista ontwikkeld, het maakt gebruik van de vertaalsoftware SYSTRAN die op meer websites wordt gebruikt. Babelfish vertaalt tussen 14 talen, het Nederlands behoort tot deze talen maar het Tsjechisch niet. De naam *Babel* komt van de Toren van Babel, die in het boek *Genesis* in het Oude Testament van de Bijbel voorkomt. Verder is de naam Babelfish afkomstig van de Babelvis, een fictief dier dat voor directe vertaling wordt gebruikt in het boek *Hitchhiker's Guide to the Galaxy* van Douglas Adams. De mensen en andere intelligente wezens plaatsen de vissen in een oor, waar zijn voedingsproces de geluidsgolven in hersengolven omzet, die tot onmiddellijke taalvertaling leiden.

*The Babel fish is small, yellow and leechlike, and probably the oddest thing in the Universe. ... The practical upshot of all this is that if you stick a Babel fish in your ear you can instantly understand anything said to you in any form of language.*⁹⁶

Afbeelding 11 Een voorbeeld – machine vertaling van Babel Fish

The screenshot shows the Yahoo! Babel Fish translation interface. At the top, there is a search bar with the text "YAHOO! BABEL FISH" and a "Search" button. Below the search bar, there are two main panels. The left panel is titled "In het Nederlands" and contains the Dutch text: "Babelfish is klein, geel en leechlike, en waarschijnlijk het meest oneven ding in het Heelal. ... Het praktische resultaat van dit alles is dat als u een Babelfish in uw oor plakt u om het even wat kunt onmiddellijk begrijpen bovengenoemd aan u in om het even welke vorm van taal". Below this text is a button that says "Search the web with this text". The right panel is titled "Translate again" and contains the English translation: "The Babel fish is small, yellow and leechlike, and probably the oddest thing in the Universe. ... The practical upshot of all this is that if you stick a Babel fish in your ear you can instantly understand anything said to you in any form of language". Below the translation is a dropdown menu set to "English to Dutch" and a "Translate" button.

⁹⁵ "Google Translation". 2009. Vertaalwerk. 2 februari 2012 <<http://vertaalwerk.jouwpagina.nl/>>.

⁹⁶ Douglas Adams, *The Hitchhiker's Guide to the Galaxy* (NY: Harmony books, 1979) 49-50.

5.3.2 Corpus-based MT

De computer maakt hierbij gebruik van een gigantische databank met bestaande vertalingen. Deze vertalingen kan hij gebruiken als voorbeeld. Hij gaat op zoek naar zinnen in de databank die zoveel mogelijk lijken op de zinnen in zijn tekst. Op basis hiervan berekent hij hoe zijn tekst waarschijnlijk in de doeltaal omgezet moet worden. Bij deze methode is de betekenis meestal goed, bijvoorbeeld ook enkele vaste uitdrukkingen maar de grammatica is niet zo goed bijvoorbeeld verbuiging en vervoeging van het Tsjechisch of woordvolgorde.⁹⁷

5.3.2.1 Google Translate

Google Translate staat bekend om de statistische methode. Google zegt over zijn eigen programma: *Wij voeren de computer miljarden woorden aan tekst, zowel eentalige teksten in de doeltaal als uitgelijnde teksten die door mensen zijn vertaald. Vervolgens passen we statistische leertechnieken toe om een vertaalmodel te maken. De evaluatie van dit onderzoek heeft zeer goede resultaten opgeleverd.*⁹⁸ Google Translate is een gratis vertaaldienst die directe vertalingen tussen 60 verschillende talen verschaft, dat wil zeggen 3540 vertaalcombinaties. Tot deze talen behoren het Tsjechisch en het Nederlands en verder ook een aantal “exotische talen” zoals het Albanees, het Servisch of het Swahili. De dienst kan woorden, zinnen en hele webpagina’s vertalen met elke combinatie van ondersteunde talen. Aangezien de vertalingen worden gegenereerd door computers, zijn ze vaak niet perfect. Hoe meer door mensen vertaalde documenten Google Translate kan analyseren in een bepaalde taal, hoe beter de kwaliteit van de vertaling wordt.

The corpus it can scan includes all the paper put out since 1957 by the EU in two dozen languages, everything the UN and its agencies have ever done in writing in six official languages, and huge amounts of other material, from the records of international tribunals to company reports and all the articles and books in bilingual form that have been put up on

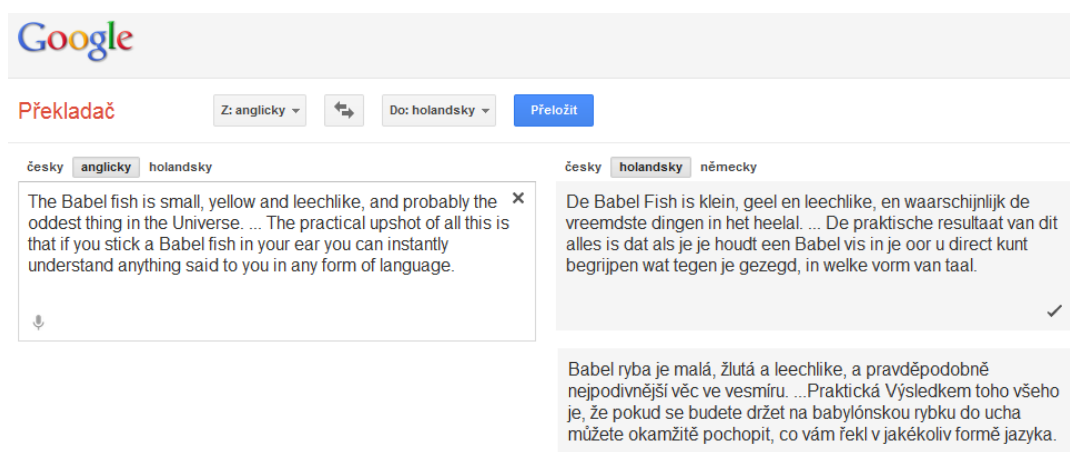
⁹⁷“Vertalen met Babel Fish: De online-tolk bij uitstek“. 6 juli 2006. Clickx. 19 februari 2012 <<http://www.clickx.nl/hintsentips/58376/vertalen-met-babel-fish/>>.

⁹⁸“Překladač Google“. 2011. Google. 20 februari 2012 <http://www.google.nl/help/faq_translation.html>.

*the web by individuals, libraries, booksellers, authors and academic departments.*⁹⁹

De nauwkeurigheid van de vertaling verschilt per taal. Als het om twee kleinere talen gaat, is de tekst eerst vertaald naar het Engels en pas daarna naar een doeltaal. Om deze reden kan het gebeuren dat een vertaling Tsjechisch-Pools op eens een Engels woord bevat. Men kan vertaalkwaliteit verbeteren door klikken op “een betere vertaling voorstellen“ en verbeteringen voor te stellen.

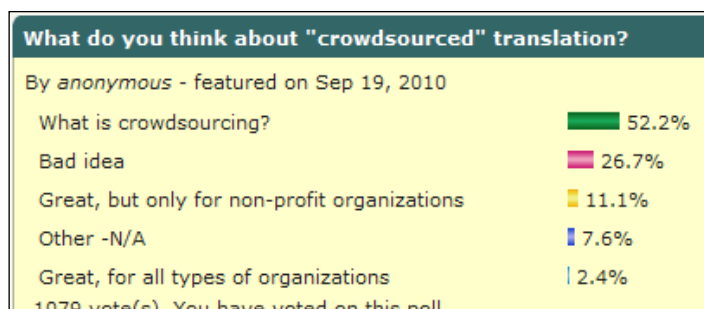
Afbeelding 12 Een voorbeeld – machine vertaling van Google Translate



5.4 Crowdsourcing – een soort “moderne vertalen”

Crowdsourcing – wat is het eigenlijk? In benedenstaande figuur 3 zijn de resultaten van de enquête verricht op ProZ te zien. Meer dan helft van de vertalers kent het begrip niet en meer dan een kwart vindt het geen goed idee.

Figuur 3 Enquête “What do you think about “crowdsourced” translation?”



⁹⁹“How Google Translate Works“. 13 september 2011. The independent. 18 maart 2012 <<http://www.independent.co.uk/life-style/gadgets-and-tech/features/how-google-translate-works-2353594.html>>.

Het begrip *crowdsourcing* is niet overal bekend. We komen dit fenomeen echter dagelijks tegen en het neemt diagonaal toe. *Crowdsourcing* is een Engelstalig neologisme, gebruikt om een recente ontwikkeling aan te duiden, waarin organisaties of personen gebruik maken van een grote groep niet vooraf gespecificeerde individuen voor beleidsvorming, innovatie of onderzoek dat wil zeggen het *gebruikmaken van een grote groep willekeurige individuen voor allerlei onderzoek*.¹⁰⁰ De term werd voor het eerst gebruikt door Jedd Howe in 2006 in zijn artikel “The Rise of Crowdsourcing”. Het begrip is een samenvoeging van de termen *outsourcing*¹⁰¹ en *crowd*, het Engelse woord voor menigte. De definitie zoals Jeff Howe hem ooit als eerste formuleerde is volgende: *Crowdsourcing is het uitbesteden van bedrijfsactiviteiten aan vrijwilligers die via het internet daartoe worden opgeroepen*.¹⁰² De Nederlander Henk van Ess kwam in september 2010 met een passende definitie: *Crowdsourcing is het bundelen van kennis van bereidwillige experts die voor niets een probleem willen oplossen en het antwoord gratis willen delen met anderen*.¹⁰³

Het fenomeen van gemeenschappelijke vertalen kan voor vertalers van belang zijn. Denk aan Wikipedia of Facebook. Duizenden vrijwilligers verlenen gratis vertalingen aan deze pagina's. Vertalingen zijn niet perfect wat geen probleem is omdat andere gebruikers fouten gewoon gaan verbeteren. Belangrijk is dat alles gratis, op vrijwilligersbasis werkt. Is dan crowdsourcing een andere soort concurrentie voor vertalers naast machinevertalen? Beide kunnen goed samenwerken – denk aan Google Translate waar gebruikers worden gevraagd om een imperfecte machinevertaling te waarderen en verbeteren.

¹⁰⁰“Crowdsourcen“ . Nederlandse Taal. 9 april 2012

<<http://www.woorden.org/woord/crowdsourcen>>.

¹⁰¹ Outsourcing of uitbesteding betekent: het laten uitvoeren van bepaalde bewerkingen of diensten of het laten produceren van bepaalde produkten door derden
<<http://www.encyclo.nl/begrip/Uitbesteding>>.

¹⁰²Jesse Gosse. 16 februari 2012. “Wat is crowdsourcing?” 27 april 2012
<<http://besocialized.nl/2011/02/16/wat-is-crowdsourcing-lees-het-hier/>>.

¹⁰³Hans Timmerman. 5 april 2012. “Het resultaat telt!” 25 april 2012
<<http://www.datacentered.nl/?p=752>>.

Enkele voorbeelden van grote crowdsourcing projecten:

- **Wikipedia**

Wikipedia werd in 2001 opgericht door de Amerikaanse ondernemer Jimmy Wales. Het doel van Wikipedia is om in veel talen een “vrije” internetencyclopedie te creëren. De Engelstalige versie is het grootste met ruim 3 900 000 artikelen. In totaal zijn er op de hele Wikipedia meer dan 21 000 000 artikelen in 284 verschillende talen te vinden. Artikelen worden door vrijwilligers geschreven, er zijn ruim 100 000 actieve leden die regelmatig bijdragen. Bij elkaar wordt Wikipedia opgebouwd door vele duizenden mensen over de hele wereld: wetenschappers, hobbyisten, studenten, experts en mensen met een goede algemene ontwikkeling. Iedereen kan aan Wikipedia bijdragen, ook anoniem. Er is geen limiet aan de bijdragen. Sommigen dragen incidenteel bij, anderen dagelijks. Alle tekst kan door iedereen bewerkt worden, dus als iemand een kort artikel begint, kunnen anderen eraan verder werken tot het een volwaardig encyclopedisch artikel is.¹⁰⁴

Wikipedia is een moderne wijze van de documentering van de menselijke kennis. Het maakt gebruik van de nieuwe mogelijkheden van www, internet en software voor samenwerking. Wikipedia is opgezet in de vorm van een *wiki* wat betekent dat elke inhoudelijke pagina door iedere willekeurige bezoeker bewerkt kan worden. *The crowdsourced nature of Wikipedia's content creation means that anyone can add falsehoods to, or vandalize, the site. However, it also enables people to easily correct such mistakes.*¹⁰⁵

- **Facebook**

Facebook bestaat sinds 4 februari 2004. Mark Zuckerberg is de oprichter en algemeen directeur. In eerste instantie was het een universiteitsproject. Alleen studenten van Harvard College konden lid worden. Nu is het derde grootste “land” ter wereld (na China en India) met 845 000 000 actieve leden.

Spaans, Duits en Frans waren de eerste talen waarin Facebook in 2008 werd vertaald. Binnen twee jaar is Facebook in 75 talen vertaald waarvan in 20 door professionele vertalers, de rest door Facebook leden. Er zijn 400 000

¹⁰⁴“Wikipedia“. 18 april 2012 <<http://nl.wikipedia.org/wiki/Wikipedia>>.

¹⁰⁵“Wikipedia“. 18 april 2012 <<http://en.wikipedia.org/wiki/Wikipedia>>.

vrijwilligers waarvan 100 000 wekelijks bijdragen.¹⁰⁶ De TAUS¹⁰⁷ gemeenschap vraagt zich af: *if 100,000 Facebook users are willing to help Facebook with translations, what will happen if that population is asked to help each other? It is easy to imagine that a community larger than the global professional translation workforce will be mobilized and do volunteer translations. ... We [TAUS] include Facebook in our 21st century outlook on translation, not for what they do today, but for the tremendous potential that the largest social networking community brings to a translation solution. We are excited to see whether they, and others, will pick up on this potential.*¹⁰⁸

- **Twitter**

Twitter is opgericht in maart 2006 in San Francisco. Twitter is een gratis internet dienst waarmee gebruikers korte berichtjes van maximaal 140 tekens publiceren. Het is een sociale netwerksite waarop men van zichzelf, zoals bij Facebook, een profiel met een foto kan maken. In 2009 vroeg Twitter zijn gebruikers om aan het vertalen van de pagina deel te nemen. De eerste talen te vertalen waren Frans, Italiaans, Duits en Spaans. Op 14 februari 2011 hebben ze *Vertaalcentrum* (Translation center) voorgesteld wat betekent dat niet alleen leden van Twitter maar alle vrijwilligers waren hartelijk uitgenodigd om de vertaling uit te voeren. Nu hebben ze 29 talen, waaronder Nederlands of Hongaars maar ook Japans, Chinees of Hindi.¹⁰⁹

- **Oxford English Dictionary (OED)**

Lang voor het ontstaan van het begrip *crowdsourcing* hebben duizenden vrijwilligers deelgenomen aan het ontstaan van de OED (vanaf 1857 tot 1928). Het ontstaansproces van de OED kan worden beschouwd als een van de vroegste voorbeelden van crowdsourcing. Er werd een oproep gedaan om vrijwilligers aan te spreken die Engelse woorden in een context zouden verzamelen. Over de periode van 70 jaar hebben de ontwerpers meer dan 6 miljoen bijdragen teruggekregen. *Prior to the OED, most English dictionaries*

¹⁰⁶Jaap Van Der Meer. 2 august 2010. "Where are Facebook, Google, IBM and Microsoft taking us?" TAUS. 22 april 2012 <<http://www.translationautomation.com/perspectives/where-are-facebook-google-ibm-and-microsoft-taking-us.html>>.

¹⁰⁷TAUS – Translation Automaton User Society <<http://www.translationautomation.com/>>.

¹⁰⁸Jaap Van Der Meer <<http://www.translationautomation.com/perspectives/where-are-facebook-google-ibm-and-microsoft-taking-us.html>>.

¹⁰⁹"Translating Twitter into more languages". 14 februari 2011. Twitter blog. 19 april 2012 <<http://blog.twitter.com/2011/02/translating-twitter-into-more-languages.html>>.

*were inconsistent, lacking in historical context and about as far from comprehensive as it was possible to be. One of the reasons the OED was able to change that was because hundreds of volunteers were assigned to particular books, copying passages illustrating word usage onto quotation slips. There were errors here and there but the project proved you could use the general public to make labour intensive tasks quicker and cheaper.*¹¹⁰

¹¹⁰Stuart Thomas. 15 september 2011. "9 examples of crowdsourcing, before 'crowdsourcing' existed". 20 maart 2012 <<http://memeburn.com/2011/09/9-examples-of-crowdsourcing-before-%E2%80%98crowdsourcing%E2%80%99-existed/>>.

II Een praktisch gedeelte

6 Enquête resultaten

De bedoeling van de enquête was om gegevens over het gebruik van vertaalprogramma's te verzamelen. De enquête was samengesteld voor vertalers tussen twee kleinere talen vooral NL-CS, CS-NL¹¹¹ en vertalers tussen een grotere en een kleinere taal EN-NL, NL-EN die met CAT tools werken. De bedoeling was om tendenties in het gebruik van vertaalprogramma's te zien. Mochten er verschillen zijn, worden ze besproken en geanalyseerd. Belangrijke punten waren de redenen om een specifieke CAT tool aan te schaffen, de opinies van vertalers over vertaalprogramma's – voordelen en nadelen, verder ook afgeronde opleiding van vertalers en hun specialisatie, gebruik van MT en hun opinie daarover enz.

In totaal hebben 21 vertalers aan de enquête deelgenomen. 5 vertalers met twee kleinere talen, 11 met een grote en een kleinere taal en 5 die beide combinaties hebben. Het was lastig om contacten met vertalers met een geschikte taalcombinatie die bovendien met CAT tools werken te verzamelen. Ten eerste heb ik via internet vertaalbureaus en vertalers gevonden die vertalingen in geschikte talen aanbieden. Ik heb echter weinig antwoord teruggekregen en een groot aantal antwoorden was negatief omdat ze niet met CAT tools werkten. Verder heb ik de enquêtes op Tsjechisch forum¹¹² en op Facebook geplaatst. Het meest respondenten heb ik op www.ProZ.com gekregen. Ik heb daar vertalers met geschikte talencombinaties gevonden en gecontacteerd. Sommigen waren behulpzaam, hebben me ook aanvullende informatie verleend en een interesse in resultaten van dit onderzoeksproject getoond.

In dit hoofdstuk worden de resultaten van de enquête verwerkt en besproken. Om enquêtes te bewerken wordt het programma Sasd LITE¹¹³ gebruikt dat statistische gegevens analyseert en vergelijkt. Sommige resultaten worden verder vergeleken met enquêtes op [ProZ.com](http://www.ProZ.com) die door vertalers samengesteld zijn en vaak meer dan duizend respondenten van de hele wereld hebben en verder ook met een onderzoek verricht door de *Jednota tlumočnicků a překladatelů* (JTP).¹¹⁴

¹¹¹Sommige vertalers hebben ook Slowaaks (in plaats van of naast Tsjechisch).

¹¹²Tsjechie.net, thuis in Tsjechië. 20 maart 2012 <<http://www.tsjechie.net/forum/index.php>>.

¹¹³Statistická Analýza Dat. 20 maart 2012 <<http://www.sasd.cz/informace/home>>.

¹¹⁴Jednota tlumočnicků a překladatelů (Vereniging van tolken en vertalers <<http://www.jtpunion.org/spip/>>.

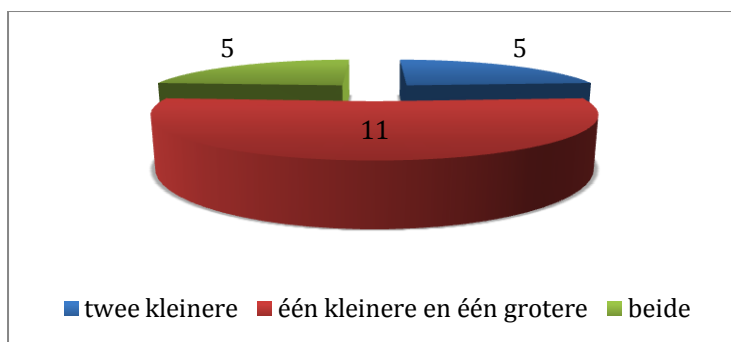
Enquête resultaten beschikbaar op de volgendepagina: <<http://utrl.ff.cuni.cz/UTRLLFF-75-version1-AnketaVyzkumTrhu.pdf>>.

Het onderzoek werd verricht door JTP in 2009 en verwerkt in 2010. Er worden 1.200 enquêtes uitgegeven (tijdens Jeronýmovy dny 2009 en via tijdschrift ToP) en 209 ingevulde enquêtes zijn teruggekomen dus 17,4 %.

6.1 Talencombinaties

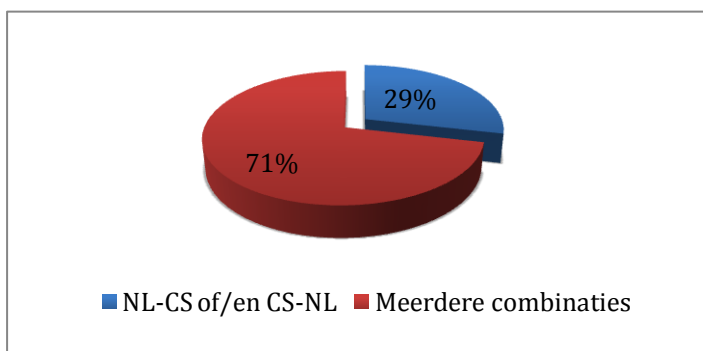
De bedoeling was om verschillen in het gebruik van vertaalprogramma's voor talencombinaties NL-EN of/en EN-NL en NL-CS of/en CS-NL te analyseren. Het blijkt echter dat een groot aantal vertalers die tussen twee kleinere talen vertalen nog andere taalcombinaties hebben, in dit geval is het 50% vertalers. De tweede combinatie was altijd Engels (4x EN-NL/NL-EN en 1x EN-CS/CS-EN) zie figuur nr 4. Een groep van vijf vertalers kan moeilijk als een representatieve steekproef beschouwd worden. Om deze reden worden alle resultaten als één geheel besproken. In geval dat er opmerkelijke verschillen tussen de groepen zullen zijn, worden ze apart besproken en geanalyseerd.

Figuur 4 Aantal vertalers



Er werd verder een onderzoek gedaan op ProZ.com om de hypothese te ondersteunen dat vertalers tussen twee kleinere talen vaak nog andere talencombinatie(s) hebben, vooral met wat grotere talen. In totaal zijn er 5 leden met talencombinatie NL-CS of/en CS-NL op ProZ.com geregistreerd waarvan iedereen meerdere talen combinaties heeft, namelijk vier vertalers hebben Engels en twee Duits plus verdere talen (Frans, Russisch, enz.). Er zijn verder 42 geregistreerde vertalers met combinatie NL-CS of/en CS-NL op ProZ.com die echter geen leden zijn. Van 42 geregistreerde vertalers hebben slechts 12 alleen de combinatie NL-CS of/en CS-NL dat wil zeggen dat 30 vertalers (71%) meerdere taalcombinaties hebben.

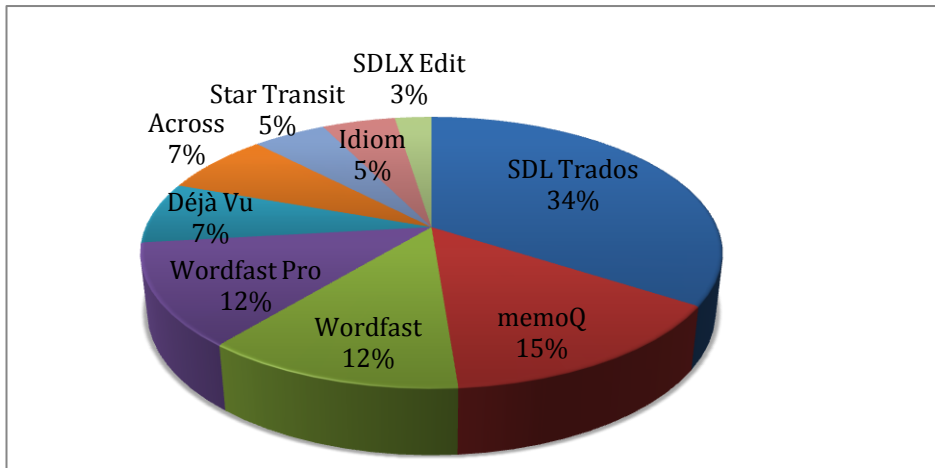
Figuur 5 Talencombinaties op ProZ.com



6.2 Vertaalprogramma's

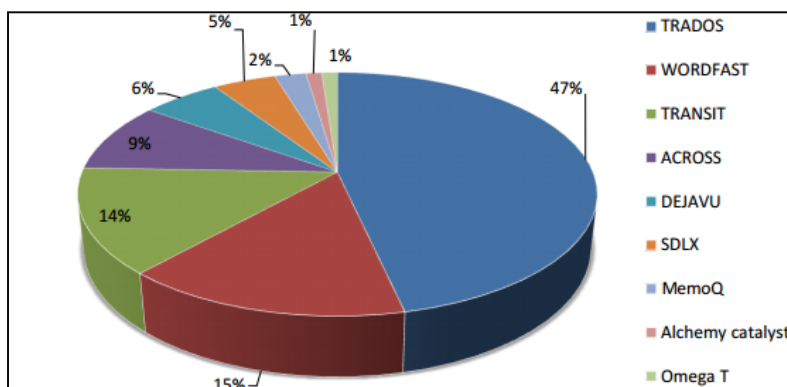
6.2.1 De meest gebruikte programma's

Figuur 6 Programma's



Het meest gebruikte programma is SDL Trados (36 %) hoewel SDL Trados het duurste programma is (zie hoofdstuk 4.2.2 Prijs van de programma's). De volgende programma's zijn er meer dan 10% percent van de vertalers aanwezig, namelijk MemoQ (15%), Wordfast en WordfastPro beide met 12%. Verder Déjà vu en Across met 7%, Star Transit en Idiom met 5% en SDLX Edit met 3%.

Figuur 7 Programma's (JTP)



De resultaten kunnen vergeleken worden met het onderzoek van de JTP. In het algemeen kunnen we zeggen dat de resultaten gelijke tendentie tonen. Trados (47%) op de eerste plaats, verder Wordfast (15%) en de rest minder dan 10%. Het grootste verschil is er bij Transit 14% vs. 5% en MemoQ 2% vs. 15% deze lijken plaatsen te hebben gewisseld. MemoQ is het jongste van programma's dat echter vlug groeit en steeds populairder wordt onder

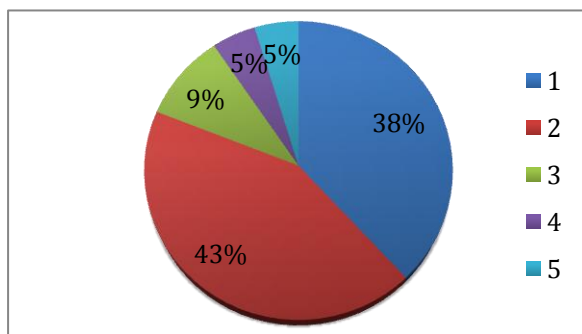
vertalers.¹¹⁵ Dat is duidelijk ook op TGB op ProZ.com waar vertalers programma's voor lagere prijzen kunnen kopen. In benedenstaande tabel nr 14 kunt u verschillende CAT tools vergelijken. In januari 2012 heeft MemoQ één keer meer licenties dan SDL Trados en zelfs vier keer meer dan Déjà Vu aangeboden. Alle zijn verkocht zelfs met een aantal vertalers op de wachtlijst. De situatie in april (23.4.2012) was gelijksoortig. Memoq heeft de meeste licenties aangeboden waarvan alle zijn verkocht, zelfs met aantal vertalers op de wachtlijst. Een van de redenen van zo'n succes kan zijn dat het een nieuw perspectief bedrijf (Kilgray) op de markt is dat een eigen product aanbiedt voor een toegankelijker prijs dan SDL Trados (zie hoofdstuk 4.2.2 Prijs van de programma's).

Tabel 14 *Verkochten licenties op ProZ*

	Januari 2012			April 2012		
	Aangeboden licenties	Verkochten licenties	Wachtlijst	Aangeboden licenties	Verkochten licenties	Wachtlijst
MemoQ	120	120	8	100	100	30
SDL Trados	60	60	35	75	60	0
Déjà vu	30	30	0	35	20	0

6.2.2 Hoeveel programma's per vertaler

Figuur 8 *Aantal programma's*



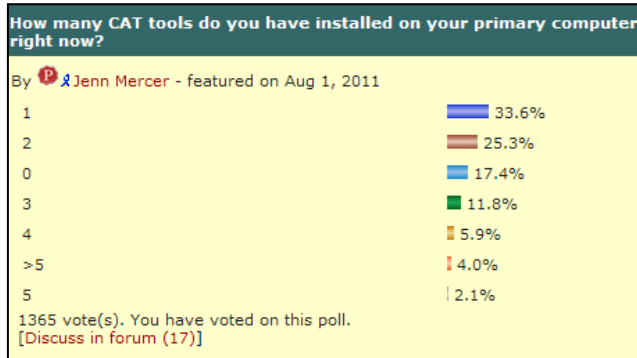
In figuur nr 8 kunt u zien hoeveel vertaalprogramma's vertalers hebben. Het meest heeft één (38%) of twee (43%). Vier en vijf programma's hebben steeds één vertaler dat wil zeggen 5%. Een vertaler met vijf programma's zegt dat nieuwe programma's leren kennen zijn/haar hobby is.

De resultaten kunnen met een enquête van ProZ.com (figuur nr 9) vergeleken worden waar ook het hoogste aantal vertalers één of twee programma's gebruiken. Verder zijn er

¹¹⁵“The price of CAT tools for freelance translators in 2012”. 30 januari 2012. YouTube.com. 20 april 2012 <<http://www.youtube.com/watch?v=dgvq2s-zymw>>.

17,4% respondenten die geen programma hebben. Vertalers die meer dan drie programma's hebben, komen niet zo vaak voor (rond 5 %) zoals in de eerste enquête.

Figuur 9 Aantal programma's (ProZ)



60% vertalers tussen 2 kleine talen heeft één programma en alleen 20% heeft er twee, niemand drie. Slechts 18% met een kleinere en een grotere taal heeft één programma en 64% heeft er twee, 18% drie. Aan welke programma's geven ze dan de voorkeur?

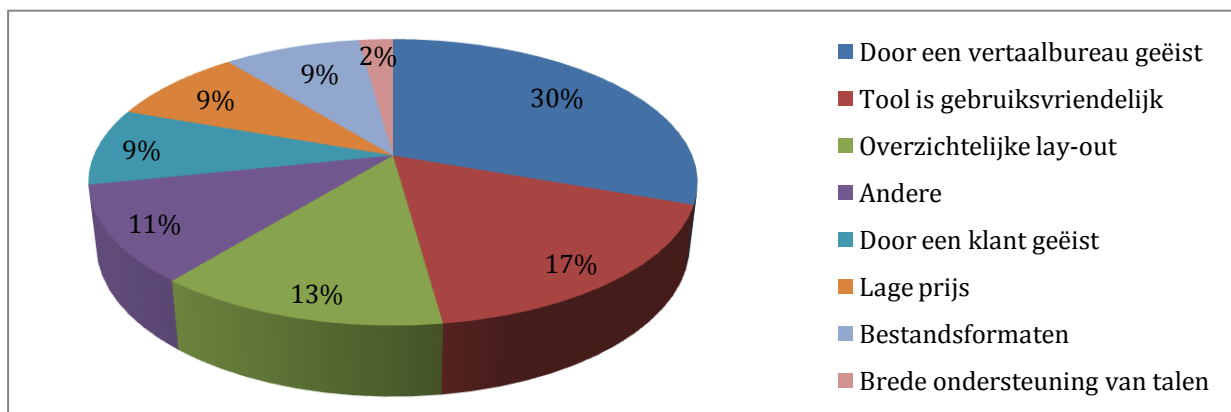
Tabel 15 Programma preferenties

	Trados	Wordfast	MemoQ	Andere
Kleine & grote (11 vertalers)	<u>10</u>	3	4	5
2 kleine + beide (10 vertalers)	4	<u>6</u>	2	5

Het is duidelijk dat vertalers tussen EN/NL de voorkeur aan Trados geven terwijl vertalers NL/CS een voorkeur aan Wordfast geven. De reden kan zijn dat vertalers NL/CS niet zo veel geld aan het programma's willen uitgeven de redenen zijn dat ze grotendeels één programma hebben en dat is vaak Wordfast, dat aanzienlijk goedkoper is dan Trados.

6.2.3 Redenen om een bepaald vertaalprogramma te kiezen

Figuur 10 Reden(en) om een CAT-tool te kiezen

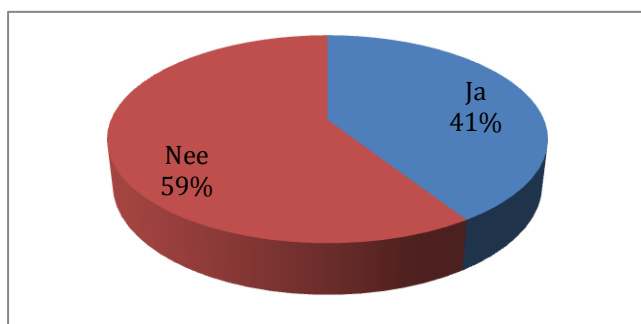


In de figuur nr 10 zijn redenen van de vertalers om een bepaalde CAT tool te kiezen te zien. De belangrijkste reden (30%) is omdat het door een vertaalbureau geëist wordt. Tweede reden (17%) is dat het programma gebruiksvriendelijk is. Op derde plaats (13%) is een overzichtelijke lay-out. Verder (11%) zijn dat andere redenen namelijk compatibiliteit met Mac (geval van Wordfast), Trados omdat het vaak gebruikt wordt, Wordfast (extra korting, lange testperiode). Met 9% zijn dat verder lage prijs, brede ondersteuning van bestandsformaten en omdat het programma door een klant geëist wordt. De laatste met 2% is de brede ondersteuning van talen.

6.3 *Is professioneel vertalen nog mogelijk zonder CAT tools*

Steeds meer vertaalbureau's verwachten van nieuwe vertalers dat ze met CAT tools werken. Zijn ze echter noodzakelijk? De enquêtevraag was volgende: *Is professioneel vertalen nog mogelijk zonder CAT tools?* Uit benedenstaande grafiek nr 11 blijkt dat beide stadpunten een groot aantal voorstanders hebben. De opinie dat vertalen zonder CAT tools niet meer kan heeft echter de overhand met 59% stemmen. De resultaten kunnen we vergelijken met een enquête die op ProZ.com werd uitgevoerd (zie figuur 12). De resultaten zijn gelijkwaardig. Meer dan helft vertalers (53,3%) denken dat kennis van CAT tools noodzakelijk voor vertalers is.

Figuur 11 *Is professioneel vertalen nog mogelijk zonder CAT tools?*



Figuur 12 *Do you think CAT tool knowledge is essential for translator?(ProZ)?*



De opinies van vertalers verschillen in dit opzicht van elkaar (zie tabel 16).

Tabel 16 *Is volgens u vertalen voor de zakelijke markt nog mogelijk zonder CAT-tools?*

	Ja	Nee
Kleine & grote (11 vertalers)	3	8
2 kleine + beide (10 vertalers)	6	4

Er is een tendentie dat vertalers tussen kleinere talen zich het werk zonder CAT tools beter kunnen voorstellen. Een van de vertalers zegt: *het is niet nuttig, het is makkelijker.* (enquête nr 14) Maar het is heel individueel, het hangt van elke vertaler/vertaalster af, of hij/zij werken met of zonder cat kiest. Er zijn enkele reacties voor:

Ik heb veel meer werk gekregen sinds ik Wordfast heb aangeschaft. Klanten willen graag consistentie en tijdsbesparing en daarmee geldbesparing. (enquête nr 20)

Nagenoeg alle vertaalbureaus vragen het. Wie niet met een cat-tool werkt, wordt ook uit de markt geprijsd.(enquête nr 4) ... *Grote opdrachten alleen met cat.* (enquête nr 5)

Het is niet eenvoudig, maar zeker wel mogelijk. Je zult alleen langer moeten werken indien je geen gebruik maakt van cat-tools. Bovendien loop je werk mis, omdat veel vertaalbureaus gebruik maken van cat-tools. (enquête nr 13)

Een van de belangrijkste redenen vóór CAT tools is zeker de coherentie van de teksten vooral bij projecten waar meerdere vertalers aan werken wat hangt samen met terminologielijsten en vertaalgeheugens. Ter tweede is het zeker tijdsbesparing en hogere productiviteit van vertalers. Ten derde is er een meer pragmatische reden die door meerdere vertalers vermeld wordt namelijk CAT tools zijn noodzakelijk omdat het door vertaalbureau's geëist wordt. Vertalers met ervaring die al jaren lang vertalen, eigen klanten hebben en niet meer via vertaalbureau's werk zoeken, kunnen zelf voorwaarden bepalen en zo bijvoorbeeld CAT tools vermijden. Bovendien kennen de klanten CAT tools vaak niet. Voor beginnende vertalers is de kennis van CAT tools echter noodzakelijk om het werk via een vertaalbureau te krijgen. Vertaalbureau's gaan dankzij CAT tools geld besparen voor terugkomende woorden en verder maken ze ook de controle van vertalers eenvoudiger (zie hoofdstuk Voordelen en nadelen van CAT programma's 3.2).

Er zijn enkele reacties tegen:

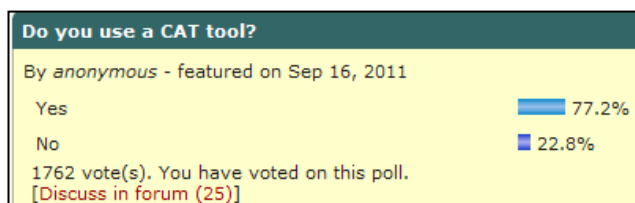
Cat tools worden misschien verlangd door veel vertaalbureaus en klanten (omdat ze denken dat de dingen dan sneller gaan), maar het produceren van een coherente en goed geschreven tekst is duidelijk gemakkelijker zonder een cat tool. je weet namelijk waar een alinea begint en eindigt en dat maakt dat je een betere tekst maakt.eigenlijk verplaatst je werk zich van vertalen wat er staat aan de linkerkant naar lezen wat het cat tool denkt dat er staat en vergelijken. tenzij het gaat om 90% matches kan ik zeggen dat ik sneller ben als ik gewoon zelf vertaal dan moet vergelijken. (enquête nr 2)

Het hangt van elke vertaler/vertaalster af, of hij/zij werken met of zonder CAT kiest. als je de literaire teksten vertaalt, dan heb je het sowieso niet nodig. of als je tolkt en daarbij doe je af en toe vertalingen, dan heb je het ook niet onvoorwaardelijk nodig. (enquête nr 14)

Een van de redenen tegen is dat vertaalprogramma's niet voor alle soorten teksten geschikt zijn. Ten tweede als iemand vertalen als een bijbaantje heeft en bovendien nog vaak verschillende soorten teksten vertaalt, kan hij/zij hoge koste voor het programma moeilijk terug winnen. Sommigen vinden vlugger om vertaling zelf te doen. Een vertaalgeheugen kan echter wel handig zijn als het om twee gelijke teksten gaan. Het nadeel is echter weer een lager tarief voor zo'n vertaling.

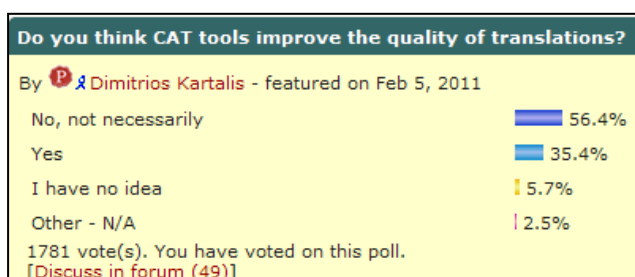
6.3.1 Vertalers en CAT tools

Figuur 13 *Do you use a CAT tool (ProZ)*



Een enquête verricht op ProZ.com toont aan dat bijna 1/4 van professioneel vertalers geen CAT tools gebruikt. En meer dan helft van vertalers (56,4%) is ervan overtuigd dat CAT tools de kwaliteit van vertalingen niet verbeteren.

Figuur 14 *Do you think CAT tools improve the quality of translations? (ProZ)*



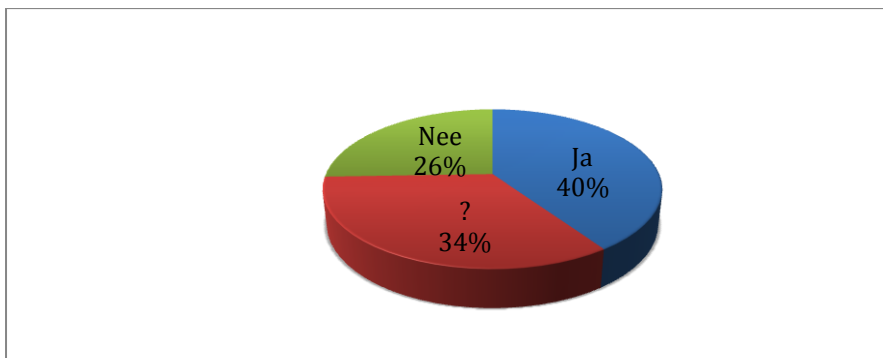
Volgens het artikel van *Active Gaming Media*¹¹⁶ dat *Are CAT tools really necessary* heet, blijkt dat CAT tools NIET noodzakelijk zijn om vertalingen van goede kwaliteit te leveren. Het programma stelt “slechts” al eerder vertaalde segmenten voor maar de vertaler moet de rest van het vertaalwerk zelf doen. Perfecte kennis van doel- en brontaal blijft noodzakelijk. Voor een persoon die een professioneel vertaler wil worden, zijn CAT tools om enkele redenen echter noodzakelijk:

- *It will help you to obtain more work. Most clients believe it is compulsory to have these tools.*
- *It will give consistency to your translations, especially when the text is long.*
- *It will allow you to work in a more organized way.*¹¹⁷

6.3.2 Vertalers NL-CS of/en CS-NL op ProZ.com

Er zijn 47 vertalers op ProZ.com die talencombinatie NL-CS of/en CS-NL hebben (5 leden en 42 geregistreerde vertalers die geen leden zijn). Alle vijf leden gebruiken CAT tools. Bij geregistreerde vertalers, ontbreken vaak aanvullende informatie dat wil zeggen 16 vertalers noemt geen programma's, er worden wel programma's vermeld bij 12 vertalers maar geen CAT tools dus we kunnen vermoeden dat ze met CAT tools niet werken of niet werkten toen het profiel opgericht was. Dat betekent dat van 47 vertalers werken met CAT tools tenminste 19 wat wil zeggen 40%.

Figuur 15 Aantal vertalers die met CAT tools werken



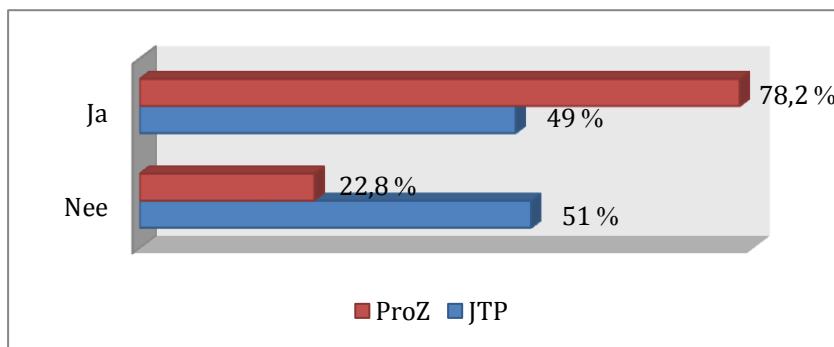
¹¹⁶Active Gaming Media is an international service company specialized in market research and marketing for video games, and is Japan's leading localization agency for video game translation and quality assurance. <<http://www.linkedin.com/company/active-gaming-media>>.

¹¹⁷“Are CAT tools really necessary?” 13 januari 2012. AGM Blog. 23 april 2012 <<http://agmbloggen.blogspot.com/2012/01/are-cat-tools-really-necessary.html>>.

6.3.3 Gebruik van CAT tools in Tsjechië

Uit het onderzoek van JTP blijkt dat 51% leden van JTP gebruiken geen CAT tools terwijl op ProZ.com is het 22,8%.¹¹⁸ In het onderzoek van JTP wordt niet besproken waarom het nummer zo hoog is.

Figuur 16 Aantal vertalers die met CAT tools werken II



Er bestaat echter een publicatie *Problems with CAT tools related to translations into Central and Eastern European languages* dat het lage aantal vertalers van Slavische talen die met CAT tools werken verklaart. Er was een onderzoek verricht tussen vertalers in Centraal en Oost Europa. Vanuit het onderzoek blijkt dat 75% vertalers beperkte kennis van programma's hebben. Sommigen hebben al over Trados of CAT tools in het algemeen wat gehoord maar ze werken met niet vertaalprogramma's. Er zijn twee redenen die door vertalers het vaakst vermeld werden:

- Prijs van CAT tools is te hoog. De salarissen in Centraal en Oost Europa zijn lager dan in West Europa. Om een CAT tool zoals Trados aan te schaffen, kan een dure investering voor een vertaler betekenen.
- Vertalers weten niet zeker wat het verschil is tussen computondersteund vertalen en machine vertalen. Ze vertrouwen de moderne techniek niet en geven ze voorkeur aan eigen werk.

Er werden ook vertaalbureau's van Centraal en Oost Europa ondervraagd. Slechts 20% vertaalbureau's maakten een gebruik van CAT tools. Bij vertaalbureau's actief op het internet was het resultaat aanzienlijk hoger (60%-70%).¹¹⁹

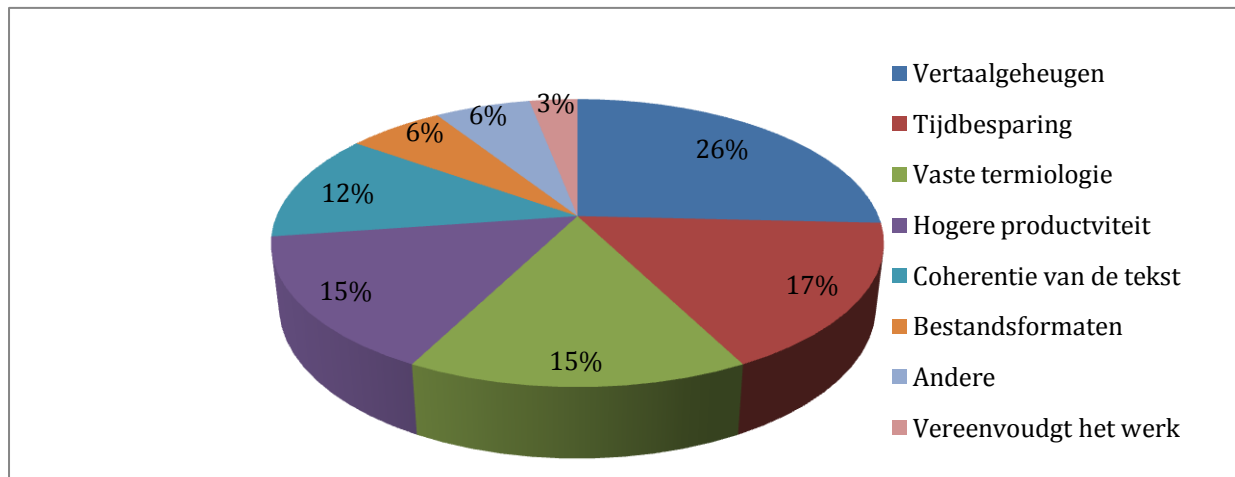
¹¹⁸Tomáš Svoboda. Januari 2011. "Anketa – průzkum překladatelského trhu". Ústav translatologie, Filozofická fakulta Univerzity Karlovy v Praze. 20 april 2011 <<http://utr1.ff.cuni.cz/UTR1FF-75-version1-AnketaVyzkumTrhu.pdf>>.

¹¹⁹Andrzej Nedoma, "Problems with CAT tools related to translations into Central and Eastern European languages", Conference *Translating and the Computer* 26 (London: Aslib, november 2004).

Sinds 2004¹²⁰ wordt de situatie op de vertaalmarkt zeker veranderd. Maar de algemene trend blijft. Vertalers actief op het internet gebruiken CAT tools vaker (78,2% op ProZ.com). Verder geeft hoge aantal vertalers van Centraal Europa namelijk Tsjechië nog steeds vaak voorkeur aan het werk zonder vertaalprogramma's (51% volgens JTP).

6.3.4 Voordelen van de programma's

Figuur 17 *Voordelen van de programma's*



In dit hoofdstuk worden de voordelen van vertaalprogramma's besproken (zie grafiek nr 17). Maximaal konden vier opties gekozen worden. Het meest vertalers beschouwen vertaalgeheugen als een voordeel (26%). Verder is het tijdbesparing (17%), vaste terminologie (15%), hogere productiviteit (15%) en coherentie van de teksten (12%). Onder 10% heeft ondersteuning van verschillende bestandsformaten (6%) en vereenvoudigen van het werk (3%). Geen van de ondervraagde vertalers beschouwt de prijs van programma's als een voordeel en iemand vindt ook het gebruik van programma's intuïtief. Verdere voordelen vermeld door vertalers (6%):

Een belangrijk voordeel is dat alle zinnen en alinea's vertaald moeten worden, het is niet mogelijk om een stuk over te slaan:

Je bent zeker dat je niets overslaat, dus dat de vertaling volledig is. Dit gebeurde vroeger regelmatig. Als een taal (en) korter is dan de andere (nl) merk je niet altijd dat er iets ontbreekt als je visueel vertaalt. (enquête nr 7)

¹²⁰Het onderzoek Problems with CAT tools related to translations into Central and Eastern European languages werd verricht in 2004.

Volgend voordeel is dat verschillende CAT tools samenwerken en gecombineerd kunnen worden. En het laatste vermeld voordeel is heel praktisch en belangrijk namelijk dat opmaak van vertaalde tekst automatisch gelijk aan brontekst is. Er zijn echter verschillen tussen preferenties van verschillende talencombinaties:

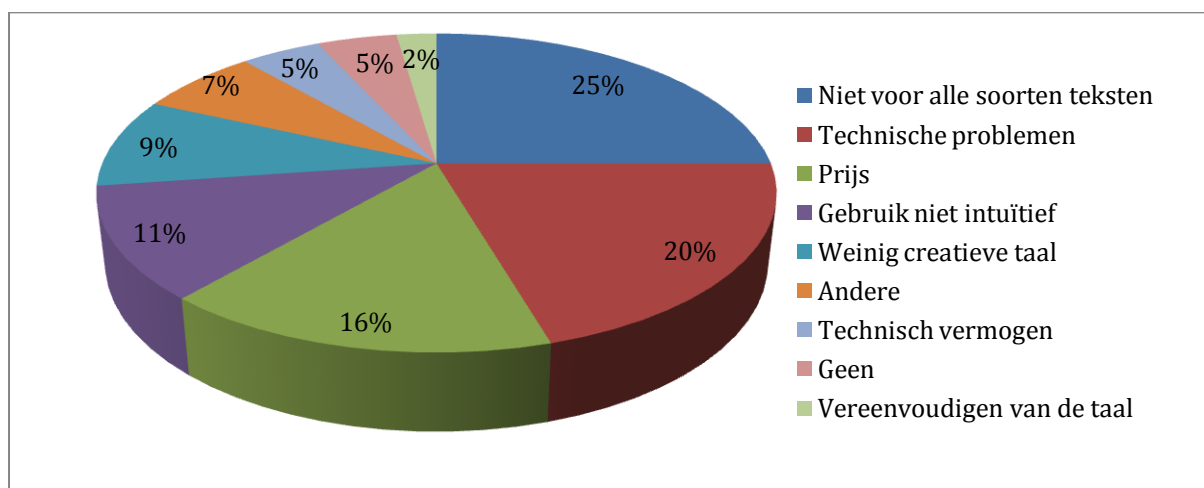
Tabel 17 Vier belangrijkste voordelen

1. Twee kleine (5 vertalers)	2. Kleinere & grotere (11 vertalers)	3. Beide (5 vertalers)
Vertaalgeheugen (4)	Vertaalgeheugen (9)	Vertaalgeheugen (4)
<u>Vaste terminologie (4)</u>	<u>Hogere productiviteit (6)</u>	Tijdbesparing (4)
Tijdbesparing (2)	Tijdbesparing (5)	<u>Hogere productiviteit (3)</u>
Coherentie van de teksten (2)	Coherentie van de teksten (5)	<u>Vaste terminologie (2)</u>

Voor alle drie groepen zijn **vertaalgeheugen** en **tijdbesparing** belangrijkste voordelen. Verder zijn er enkele verschillen. Voor de groep 1 is de **vaste terminologie** heel belangrijk maar bij de groep 2 behoort het niet tot vier belangrijkste voordelen.¹²¹ Het tweede verschil is dat de groep 2 **hogere productiviteit** noemt maar bij de groep 1 behoort het niet tot vier belangrijkste voordelen. De groep 3 staat er tussen omdat ze alle twee gemeenschappelijke punten noemt maar ook beide afwijkende.

6.3.5 Nadelen van de programma's

Figuur 18 Nadelen van de programma's



In dit hoofdstuk worden nadelen van vertaalprogramma's vermeld (zie figuur 18). Vertalers konden maximaal vier opties kiezen. Het grootste nadeel is dat CAT tools niet voor alle

¹²¹ Voor meer uitleg zie hoofdstuk 6.4 Glossarium vs. vertaalgeheugen.

soorten teksten geschikt zijn (25%). Verder zijn dat technische problemen (21%), prijs van het programma (16%), het gebruik van het programma is niet intuïtief (12%) en weinig creatieve taal 9%. 5% vertalers zeggen dat weinig creatieve taal een nadeel is en 2% vertalersnoemen vereenvoudigen van de taal.

Verder ook 5% zeggen dat ze zo geen punten kunnen opnoemen. In het algemeen wordt er minder nadelen dan voordelen genoemd (67 voordelen en 44 nadelen). Het betekent dat vertalers eerder met de programma's tevreden zijn. Slechts in twee gevallen hebben ze alle vier negatieve punten genoemd.

Verdere nadelen die vermeld worden zijn de volgende (7%):

Vermits NL en EN een omgekeerde zinsbouw hebben, krijg je problemen wanneer zinnen in segmenten verdeeld worden en de segmenten na vertaling niet meer met elkaar overeenkomen. nl: klemtoon op het begin van de zin, en klemtoon op het einde. (enquête nr 7)

Veel agentschappen gebruiken deze om de vertaler onder druk te zetten haar/zijn prijzen te verlagen, wat leidt tot de situatie dat de vertaler een investering moet doen en dan als beloning minder verdient. de kwaliteit van de vertaalgeheugens van die agentschappen is bijna altijd van een droevige kwaliteit dat ze onbruikbaar zijn en fouten veroorzaken in plaats van voorkomen. (enquête nr 18)

Beide punten werden al vroeger besproken (zie hoofdstuk 2.1.3 voor verschillen tussen talen en het hoofdstuk 3.2.2 voor nadelen van vertaalprogramma's).

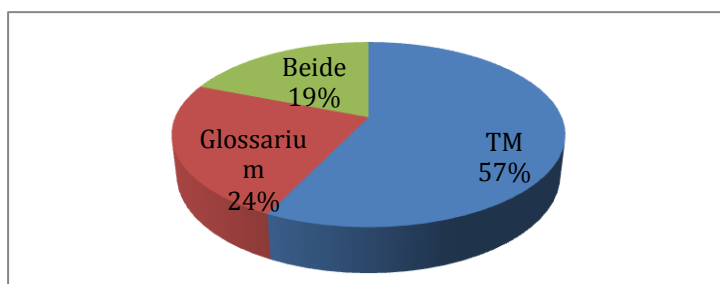
Er zijn echter verschillen tussen preferenties van verschillende talencombinaties (zie tabel 18):

Tabel 18 Vier belangrijkste nadelen

1. Twee kleine + beide (10 vertalers)	2. Kleinere & grotere (11 vertalers)
Niet voor alle soorten teksten (8)	Prijs van het programma (5)
Technische problemen (4)	Technische problemen (5)
Prijs van het programma (2)	Niet voor alle soorten teksten (3)
Technisch vermogen noodzakelijk (2)	Weinig creatieve taal (3)

6.4 Glossarium vs. vertaalgeheugen – wat belangrijker

Tabel 19 Glossarium vs. vertaalgeheugen



De volgende vraag was wat vertalers belangrijker vinden en wat ze vaker gebruiken – glossarium, vertaalgeheugen (MT) of beide gelijkwaardig. 57% zeggen het vertaalgeheugen, 24% glossarium en 19% zeggen dat beide even belangrijk zijn. Er zijn enkele redenen die in de enquêtes vermeld worden:

Vertalers met een kleinere en een grotere taal:

Vertaalgeheugen voor de context van de vertaalde woorden/zinnen. (enquête nr 15)

Vertaalgeheugen omdat glossaria meer werk vergen en voor sommige klanten overbodig zijn. (enquête nr 4)

Vertaalgeheugen. Ik gebruik het glossarium alleen als ik het geleverd krijg van een klant, zelf maak ik geen glossaria aan. (enquête nr 9)

Vertalers met beide combinaties:

TM. Het is eenvoudiger. Vind ik. (enquête nr 14)

Glossarium ... Een TM dient voor mij meer 'ter inspiratie' ... in het geval van een gedeelde opdracht gaat Glossary boven TM. Bij een opdracht die ik zelf doe, kan ik meer op mijn persoonlijke TM vertrouwen, omdat ik precies weet waarom ik een bepaalde beslissing heb genomen. (enquête nr 12)

Twee vertalers vinden vertaalgeheugen belangrijker omdat het context van de te vertalen woorden en zinnen toont. Drie vertalers gebruiken vertaalgeheugen vaker omdat het makkelijker te creëren is. Twee vertalers vinden glossaria lastig om op te zetten en importeren en ze gebruiken glossaria alleen als ze ze van de klanten geleverd krijgen. Verder blijkt dat glossaria belangrijker zijn vooral bij opdrachten waaraan meerdere vertalers samenwerken. Het zoeken in eerder vertaalde losse documenten is daardoor niet meer nodig (volgens 2

vertalers). En 2 vertalers vertrouwen liever hun eigen vertaalgeheugen. Ze gebruiken een gedeeld vertaalgeheugen alleen om ideeën op te doen.

Opinies van vertalers in dit opzicht verschillen zich van elkaar (zie tabel 19).

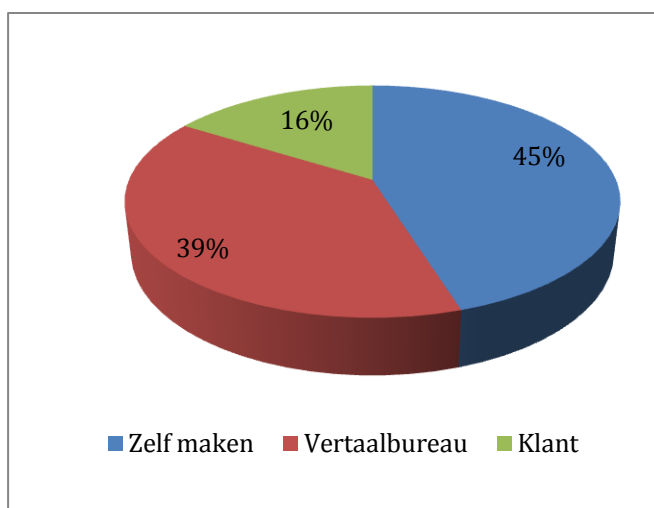
Tabel 19 *Wat vindt u belangrijker / Wat gebruikt u vaker – glossarium of vertaalgeheugen?*

	Glossarium	Vertaalgeheugen	Beide even belangrijk
1. Twee kleine (5 vertalers)	<u>3</u>	1	1
2. Kleine & grote (11 vertalers)	1	<u>8</u>	2

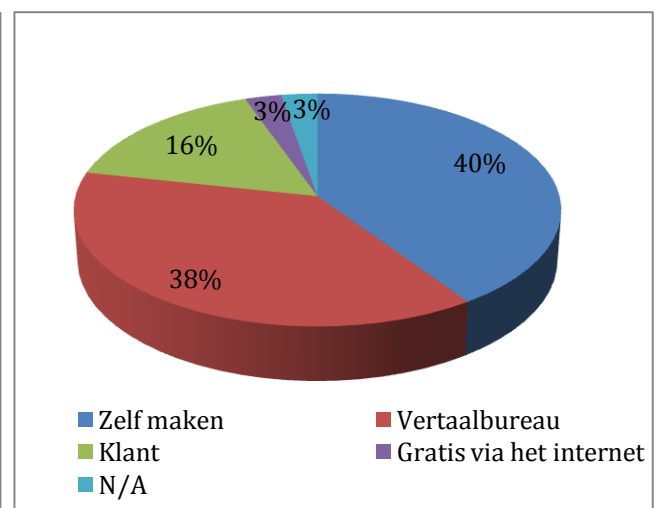
Vanuit het onderzoek blijkt dat groep 1 het glossarium belangrijker vindt en groep 2 het vertaalgeheugen. Volgens het hoofdstuk 6.3.4 (Voordelen van de programma's) wordt het al duidelijk dat groep 1 meer waarde hecht aan glossarium dan de groep 2. Een mogelijke verklaring is dat voor de talencombinatie van kleinere en grotere taal namelijk NL/EN bestaat er genoeg woordenboeken van goede kwaliteiten en ook gespecialiseerde woordenboeken die op bepaald gebied gericht zijn. Voor de talencombinatie van twee kleinere talen, namelijk NL/CS bestaat er echter niet voldoende woordenboek en gespecialiseerde woordenboeken zijn er helemaal niet. De beschikbare woordenboeken bevatten bovendien ouderwetse uitdrukkingen en zelfs ook fouten. Voor een vertaler is het daarom lastiger om een terminologie van goede kwaliteit aan te schaffen.

6.4.1 Mogelijkheden om glossarium en vertaalgeheugen aan te schaffen

Figuur 19 *Glossarium*



Figuur 20 *Vertaalgeheugen*



Uit de bovenstaande grafieken blijkt dat mogelijkheden om vertaalgeheugen en glossarium om aan te schaffen min of meer overeenkomen. Het was mogelijk om maximaal drie opties te

kiezen. Vertalers gaan het vaakst glossaria en vertaalgeheugens zelf maken, verder krijgen ze hem van vertaalbureau 's of van klanten. Het is ook mogelijk om vertaalgeheugen via het internet gratis te krijgen (3%) en één vertaler gebruikt vertaalgeheugen zelden:

*n/a voor mij (klant geeft het soms, maar is meer dan tevreden over mijn product zonder tm
blijkbaar. (enquête nr 14)*

6.5 Competenties van de vertaler van vandaag

Volgende vraag die beantwoordt moet worden is: *Welke competenties zijn noodzakelijk om tegenwoordig vertaler te worden?* Er worden enkele opinies van de ondervraagde vertalers vermeld.

Vertalers tussen twee kleinere talen:

Kennis van internet en sociale media. Technische kennis (IT kennis) en brede interesse (enquête nr13) ... specialisme (enquête nr 1)

Vertalers met een kleinere en een grotere taal:

Bereidheid om erg veel tijd te investeren in elke tekst, ook perfectionisme: veel opzoekwerk en de tekst niet afleveren voor die uitstekend is. Je houden aan de eindtermijn. Bereid zijn voortdurend bij te leren. (enquête nr 7)

Een zekere computervaardigheid, niet alleen om met cat-tools te kunnen werken en de bijbehorende technische problemen op te kunnen lossen. (enquête nr 19) ... snel informatie opzoeken (vb. parallelle teksten), vlotheid met software, algemene brede kennis. (enquête nr 6)

Vertalers met beide combinaties:

Computervaardigheden (je moet om kunnen gaan met ms office, cat-tools en zeer behendig zijn in het zoeken van informatie op het internet) en een goede kennis van de cultuur van de doeltaal. (enquête nr 12)

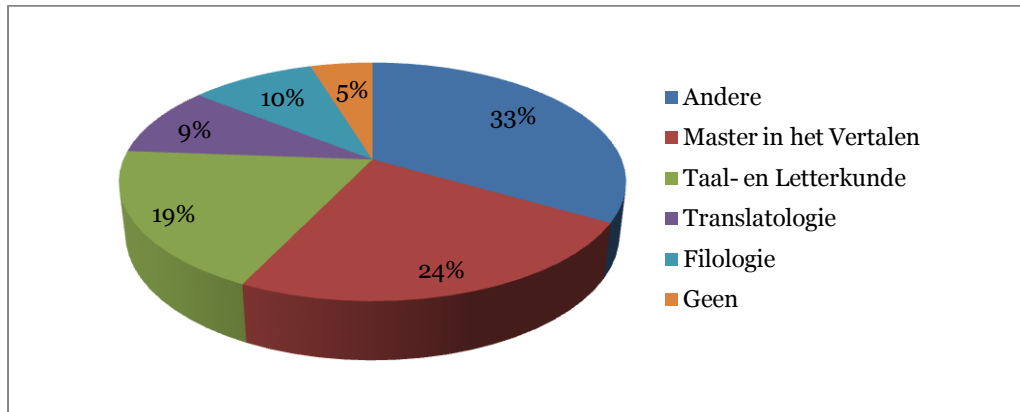
Opleiding, ervaring, ook buitenlandervaringen (wonen, werken in het land van de brontaal), flexibiliteit, goede marketing. (enquête nr 20)

Algemene technische, juridische en economische/commerciële kennis, goede praktische kennis van ms word, excel en powerpoint, communicatieve vaardigheden bij onderhandelen met klanten. (enquête nr 8)

Perfekte kennis van doel- en brontaal en beide culturen zijn natuurlijk onvermijdelijk. Verder zijn dat computervaardigheden, algemene kennis, brede interesse, specialisatie, communicatieve vaardigheden en onderhandelen. Een vertaler moet bereid zijn om voortdurend te leren en ontwikkelen. Wat eigenschappen betreft is dat zeker perfectionisme, geduld, flexibiliteit, betrouwbaarheid, klantvriendelijkheid en anderen.

6.6 Opleiding

Figuur 21 Opleiding

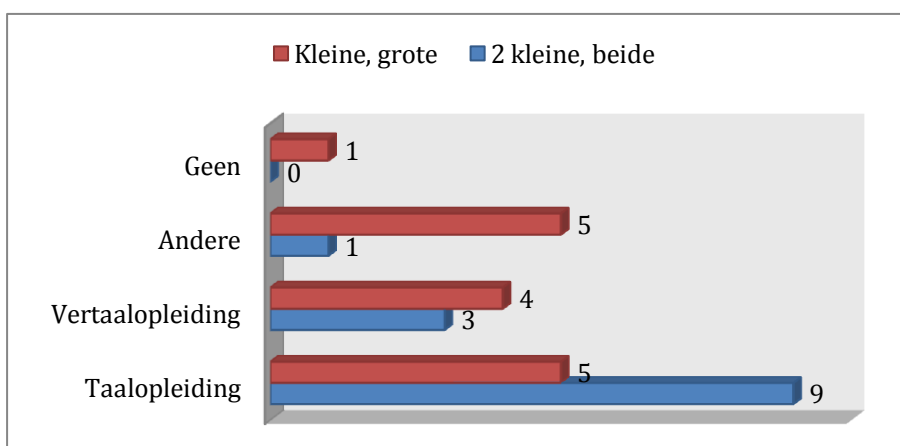


Het is opvallend dat een groot aantal vertalers geen taalopleiding heeft (38 %). Het betekent dat 62% van de vertalers een taalopleiding heeft waarvan slechts 33% vertalers Master in het Vertalen of Translatologie gestudeerd hebben.

Tabel 20 Opleiding van de vertalers

	1. Twee kleine talen (5 vertalers)	2. Kleinere + grotere (11 vertalers)	3. Beide (5 vertalers)
Taalopleiding	4	5	5
Vertaalopleiding	2	4	1
Andere	1	5	0
Geen	0	1	0

Figuur 22 Opleiding van de vertalers

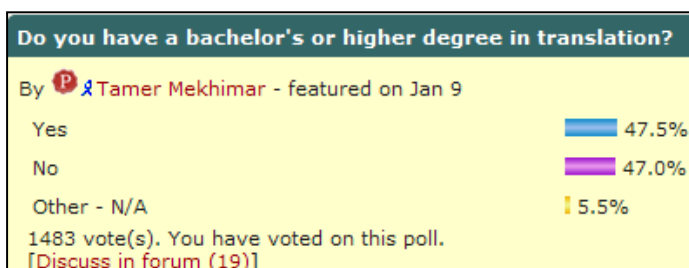


Het valt op dat in de groepen 1 en 3 waar alle vertalers talencombinatie NL-CS of/en CS-NL hebben 9 van 10 vertalers een taalopleiding hebben waarvan 3 een vertaalopleiding. Slechts één vertaler heeft een andere opleiding afgerond namelijk medicijnen. Bij vertalers met de combinatie NL-EN en/of EN-NL heeft slechts 5 van 11 vertalers talengestudeerd waarvan 4 een vertaalopleiding hebben. De volgende 5 vertalers hebben een andere opleiding afgerond namelijk politieke & sociale wetenschappen, docent Engels, master nieuwe media en digitale cultuur, fine art en filosofie. Een vertaler heeft verder geen universiteitsopleiding afgerond.

Een reden om deze resultaten te verklaren namelijk dat 90% vertalers NL-CS of/en CS-NL taalopleiding hebben en bij vertalers NL-EN en/of EN-NL is het slechts 45,5% kan zijn dat Nederlands en Tsjechisch twee kleine talen zijn en men kan moeilijk beide talen onder de knie te krijgen zonder taalopleiding. Bij een grotere en een kleinere taal in dit geval Nederlands en Engels is de situatie wat anders. In Nederland wordt Engels op school gegeven en men is vaak in contact met het Engels. Om deze reden komt voor dat men andere opleiding afrondt en pas daarna begint met het vertalen al met een bepaalde specialisatie.

Om de resultaten te vergelijken worden nog enquêtes betreffende opleiding van vertalers van ProZ.com bekijken. Het blijkt dat bijna 50% vertalers geen vertaalopleiding hebben (zie figuur nr 23):

Figuur 23 *Do you have bachelor's or higher degree in translation?(ProZ)*

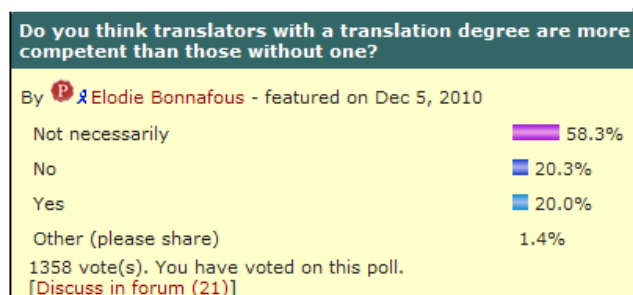


Er zijn nog twee interessante enquêtes betreffende de opleiding van vertalers. De eerste vraag is volgende: *Kan vertaalopleiding toekomstige vertalers op de praktijk voldoende voorbereiden?* (zie grafiek nr 24). Slechts 3,2% zeggen dat wel. Volgens 37,1% gedeeltelijk en volgens 45,8% helemaal niet. De tweede vraag is volgende (zie grafiek nr 25): *Zijn vertalers met een vertaalopleiding meer competent dan andere vertalers?* Slechts 20% zeggen dat wel, 20,3% dat niet en 58,3% denken dat het zo niet hoeft te zijn. Het blijkt dat een vertaalopleiding niet onvermijdelijk nodig is om een vertaler te worden.

Figuur 24 *Do universities prepare translators / interpreter sufficiently for the life in the industry?*



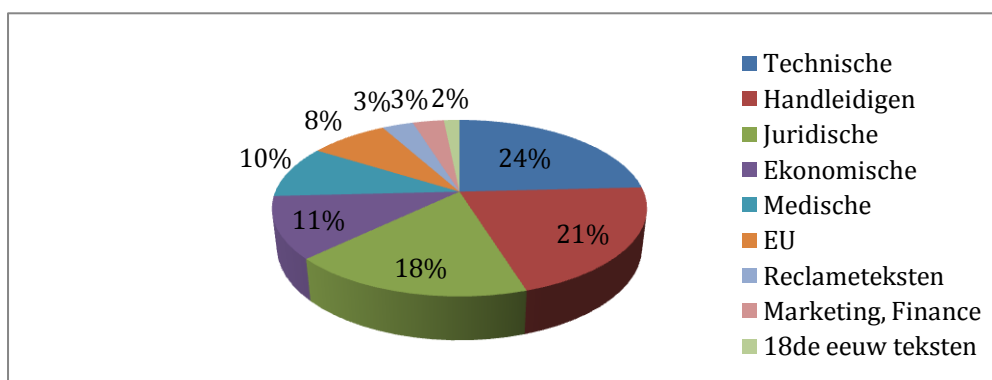
Figuur 25 *Are translators with a translation degree more competent than those without one? (ProZ)*



6.7 Soorten te vertalen testen– specialisatie van vertalers

6.7.1 Soorten te vertalen teksten

Figuur 26 *Soorten teksten*



In bovenstaande figuur 26 zijn soorten teksten te zien waarmee vertalers bezig zijn. De meest frequente zijn technische teksten (24%) gevolgd door handleidingen (21%). Verder zijn het juridische teksten (18%), economische (11%) en medische teksten (10%). Minder dan 10% van de teksten betreffen de EU (8%), reclame (3%), marketing en financiën (3%) en zelfs, met 2%, achttiende-eeuwse teksten.

Tabel 21 *Soorten teksten*

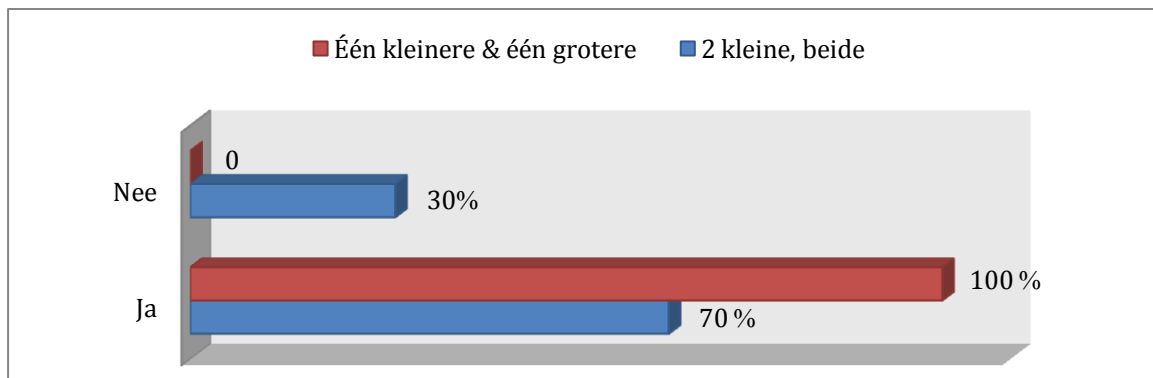
1. Twee kleine (5 vertalers)	2. Kleinere & grotere (11 vertalers)	3. Beide (5 vertalers)
Juridische (3)	Technische (9)	Technische (4)
Handleidingen (2)	Handleidingen (7)	Handleidingen (4)
Economische (2)	Juridische (4)	Juridische (4)
Technische (2)	Economische (3)	Economische (3)

In de bovenstaande tabel 21 kunt u de meest voorkomende typen te vertalen teksten zien. Het zijn in alle de drie gevallen **economische teksten, handleidingen, juridische en technische teksten**. Ze hebben slechts een beetje verschillende rangschikking, bij grotere talen komen technische teksten vaker voor.

6.7.2 Specialisatie

Is het mogelijk om zich te specialiseren? Zijn er genoeg vakteksten? Vooral voor kleinere talen kan het een probleem zijn. In onderstaande grafiek 27 zijn de enquêteresultaten te zien. Alle vertalers met één kleinere & één grotere taal zeggen dat specialisatie mogelijk is. Aan de andere kant zegt 30 % van de vertalers met twee kleine talen echter dat het moeilijk is om zich te specialiseren.

Figuur 27 Specialisatie



Enkele reacties van de vertalers met twee kleine:

Een aantal jaren krijg je een specialisatie. (enquête nr 17) ... uni opleiding is het beste. (enquête nr 1)

Enkele reacties van de vertalers met één kleinere & één grotere taal:

Ik specialiseer me nu al 15 jaar in medische teksten en teksten met betrekking tot computers. (enquête nr 18)

In mijn vakgebied is er in ieder geval voldoende tekst te vinden. (enquête nr 9)

Hangt uiteraard af van de gekozen specialisatie, maar voor IT, juridische teksten, medische teksten is er zeker genoeg werk. (enquête nr 19)

Er zijn voorbeelden genoeg op internet om bij te leren. Als je voortdurend bereid bent werk aan te nemen, geraak je bekend als een expert. (enquête nr 7)

Het hangt er natuurlijk van af welk tekstgenre je kiest. Binnen het medische vakgebied is er een overvloed aan teksten, maar ik kan me inbeelden dat dat in andere vakgebieden niet het geval is. (enquête nr 6)

Enkele negatieve opinies van de vertalers met beide combinaties:

Er zijn nog steeds te weinig vertalers in deze talencombinatie om zich te specialiseren. (enquête nr 14)

a) Ja - alleen in de “grote talen” b) Nee - in onze “kleine” taalcombinaties. Waarom? - niet genoeg brood, concurrentie van het Engels. (enquête nr 8)

Ja voor grote talen (Engels bv.), omdat er veel meer aanbod is. Nee voor kleine talen (Tsjechisch, Slowaaks), omdat het aanbod veel beperkter is. (enquête nr 12)

Enkele positieve opinies van de vertalers met beide combinaties:

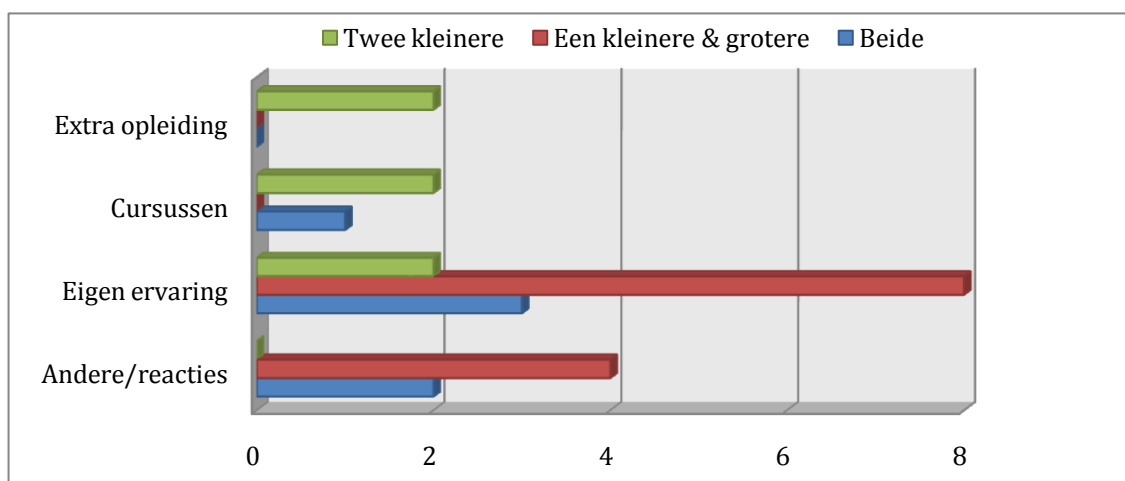
Hoewel het Nederlands een “kleinere taal” is, is het aantal vertalingen ruim wegens actieve handelscontacten tussen de landen in kwestie. (enquête nr 21)

Ja, waarom niet? Ik vertaal veel handleidingen, die zijn er altijd wel. (enquête nr 20)

Het blijkt dat het aantal vertalingen afhangt van het gekozen vakgebied. Enkele soorten teksten zoals medische, handelscontracten of handleidingen worden vaker gevraagd dan anderen. Verder blijkt uit het onderzoek dat vertalers tussen kleinere talen het moeilijker vinden om zich te specialiseren en ervan te kunnen leven omdat er concurrentie van grotere talen is en het aanbod beperkter is. Een volgend argument is dat er in deze talencombinaties niet genoeg vertalers zijn om zich te specialiseren. De opinies van de vertalers verschillen echter op dit gebied van elkaar.

6.7.3 Hoe gespecialiseerd

Figuur 28 Hoe gespecialiseerd



Gaan vertalers zich nog verder specialiseren? En hoe? Uit het onderzoek blijkt dat vertalers met kleinere talen zich vaker bijscholieren (namelijk extra opleiding of cursussen) dan vertalers met één grotere taal die allemaal de voorkeur geven aan eigen ervaringen en zelfstudie. Dit is weer verbonden met de opleiding van vertalers (zie hoofdstuk 6.6) namelijk 5 van 11 vertalers hebben een andere opleiding dan talen afgerond dus hebben ze eigenlijk al een voorkennis van een bepaald vak die ontbreekt aan vertalers die talen hebben gestudeerd.

Enkele reacties van de vertalers met een kleinere en een grotere taal:

Ik probeer niet te overdrijven. Ik doe niets te gespecialiseerd, maar doe wel graag alles met humane wetenschappen omdat me dat interesseert en dat ik daar dan ook veel over weet. (enquête nr 2)

Zelfstudie. (enquête nr 18) ... *Constant study (books, magazines, professional associations etc), and by working on my own Dutch-English dictionary for translators in all of my free time.* (enquête nr 11)

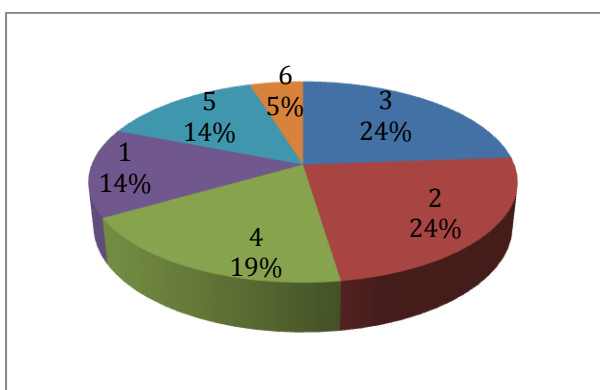
Enkele reacties van de vertalers met beide combinaties:

Veel teksten in dat specialisme [heb ik] vertaald en zo kennis opgedaan. (enquête nr 20)

Uit het onderzoek blijkt dat constante zelfstudie en een brede interesse noodzakelijk zijn om vertaler te worden. Als een vertaler jaren op een bepaald vakgebied bezig is, beschouwt hij/zij zich een expert op het gebied.

6.7.4 Aantal vakgebieden

Figuur 29 Aantal vakgebieden

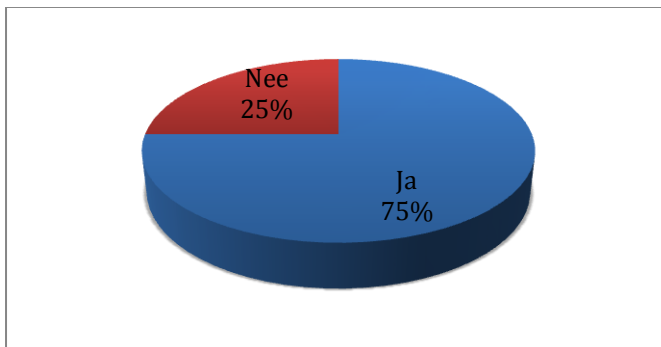


In de grafiek nr 29 kunt u zien met hoeveel verschillende soorten teksten vertalers meestal bezig zijn. Twee en drie vakgebieden zijn op de eerste plaats beide met 24% wat betekent bijna de helft van de vertalers. Boven de 10% zijn dat vier vakgebieden met 19% en één en vijf vakgebieden beide met 14%. Ten laatste is zes vakgebieden met 5%.

6.7.5 Verschillende soorten te vertalen teksten– nut van CAT tools

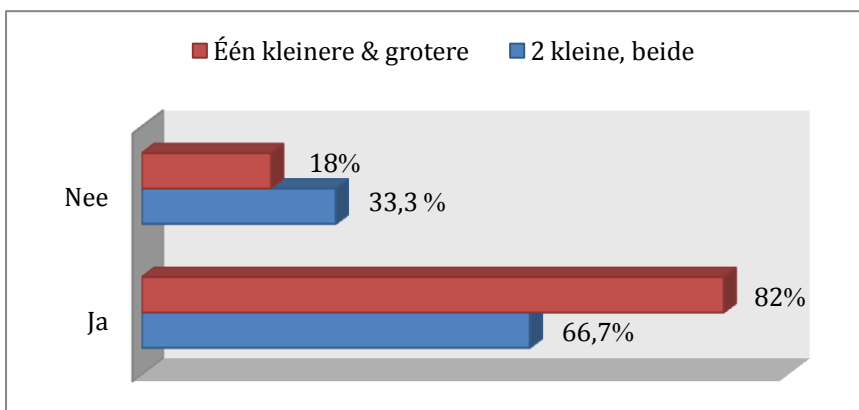
De volgende vraag luidde: *Als u vaak verschillende soorten teksten vertaalt, zijn CAT tools dan voor u nuttig?* De antwoorden zijn als volgt: één vierde vertalers zegt dat CAT tools in zo'n geval niet nuttig zijn (grafiek nr 30).

Figuur 30 *Zijn CAT tools nuttig?*



In grafiek nr 31 zijn de resultaten in twee groepen verdeeld. Uit het onderzoek blijkt dat vertalers tussen één kleinere en één grotere taal CAT tools vaker benutten namelijk 82%, in het geval van de tweede groep is het 66,7%. Enkele opinies van de vertalers kunnen de resultaten verklaren.

Figuur 31 *Zijn CAT tools nuttig? II*



Enkele opinies van de vertalers met twee kleine:

Ja, het duurt alleen wat langer voordat het product zichzelf terugbetaalt. (enquête nr 13)

Nee, weinig gebruik. (enquête nr 17)

Enkele opinies van de vertalers met een kleinere en een grotere taal:

De teksten die ik vertaal zijn veel te divers. CAT-tools zijn voor mij tijdverspilling, maar als de klant het echt wil, dan gebruik ik ze. Ik reken dan echter wel een supplement aan. Gewoonlijk leidt dat er toe dat de klant het gebruik van CAT-tools plotseling niet meer belangrijk vindt. (enquête nr 18)

Hangt van de omstandigheden af, maar een eigen vertaalgeheugen is altijd nuttig omdat je soms termen of uitdrukkingen tegenkomt die je vroeger al hebt opgezocht. (enquête nr 19) ... Omdat toch vaak weer dezelfde terminologie terugkomt en ik die dan snel terug kan vinden. (enquête nr 9)

Soms, het hangt van het soort tekst af. Een vragenlijst ... gaat dan beter vooruit met een cat, maar voor een mooie tekst is het eerder iets dat veel tijd in beslag neemt zonder veel resultaat ... elke tekst moet aangenaam zijn om te lezen. (enquête nr 2)

Tijdens het vertalen hoef je niet naar de opmaak etc. te kijken, dat doet de cat tool voor je. (enquête nr 10)

Wordfats professional en sdlx editcat-tools zijn niet alleen handig voor terminologie en tm's, maar ook omdat ze je dwingen systematisch te werk te gaan, de opmaak van documenten overnemen en ze er voor zorgen dat je geen stukken tekst mist. (enquête nr 15)

Enkele opinies van de vertalers met beide combinaties:

Nee, te veel uiteenlopende soorten teksten, vaak originelen op papier of in grafische vorm (pdf, jpg). (enquête nr 8)

Het is niet nuttig, het is makkelijker. En vrij vaak een vereiste van de klant / vertaalbureau. (enquête nr 14)

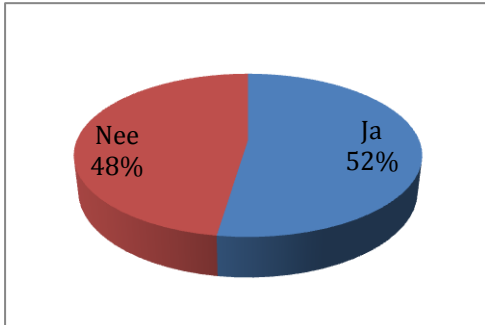
Ik gebruik de cat-tools alleen voor sommige vertaalklussen. (enquête nr 21)

Het werd al vroeger besproken dat CAT tools slechts voor enkele typen teksten geschikt zijn (zeker niet voor literaire teksten.). Soms worden ze alleen gebruikt omdat het door een vertaalbureau vereist wordt. Een zeker voordeel is dat CAT-tools een vertrouwde werkomgeving bieden, waarin alles snel gevonden kan worden. Het helpt om snel en efficiënt te werken, de coherentie van de tekst en lay-out blijven behouden, verder zijn er ook geen ontbrekende zinnen.

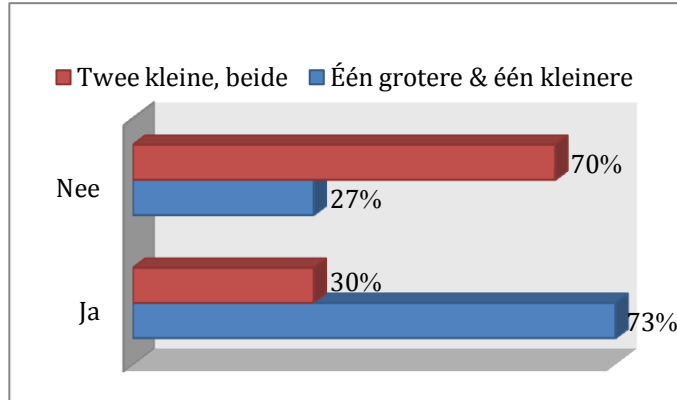
6.8 Lokalisatie

6.8.1 Vertalers en lokalisatie

Figuur 32 Lokalisatie



Figuur 33 Lokalisatie II



Meer dan helft vertalers namelijk 52% zijn ook met lokalisatie bezig. In de grafiek nr 33 kunt u zien wat voor verschillen zijn er tussen vertalers. Met lokalisatie is 73% vertalers tussen één grotere en één kleinere taal bezig maar slechts 30% vertalers tussen twee kleine talen en vertalers die beide combinaties hebben. Engels is een belangrijke taal en steeds meer materialen worden naar/vanuit het Engels vertaald daarom is de vraag hoog, dat wil zeggen aanzienlijk hoger dan tussen twee kleinere talen.

Vanuit het onderzoek blijkt dat vertalers met twee kleinere talen en beide combinaties geen gespecialiseerde programma's voor lokalisatie gebruiken:

Geen extra programma hiervoor. (enquête nr 13) ... *Geen specifieke programma's.* (enquête nr 21) ... *Ms excel + Ms word* (enquête nr 8)

Vertalers met een kleinere en een grotere taal gebruiken wel gespecialiseerde programma's voor lokalisatie. Programma's verschillen zich echter van elkaar, er is geen enkele programma dat meer gebruikers heeft:

Across, Transit (enquête nr 10) ... *Notepad, Dreamweaver, Expression Web* (enquête nr 18) ... *Passolo, Catscradle, Webbudget* (enquête nr 6) ... *Wordfats Professional en sdlx edit* (enquête nr 15) ... *Trados* (enquête nr 5) ... *Tageditor* (enquête nr 19)

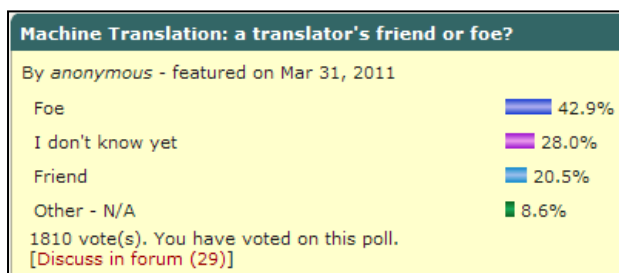
Figuur 34 Are you working in the localisation industry?(ProZ)



De resultaten kunnen verder met een enquête verricht op ProZ.com vergeleken worden. Er is 56% vertalers die niet met lokalisatie bezig zijn, daarvan zou 17% dat graag willen proberen. Alleen voor 6,2% vertalers betekent lokalisatie 81% - 100% van hun activiteiten. Het blijkt dat lokalisatie voor vertalers een aanvullend onderdeel van hun werk betekent.

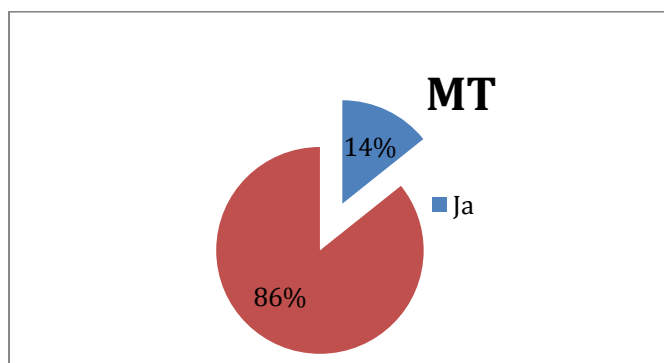
6.9 MT “Friend or foe?”

Figuur 35 Machine Translation: a translator’s friend or foe?(ProZ)



Over machine vertalen wordt veel gesproken maar wat denken vertalers daarover? Is machine vertalen hun “friend or foe”? Uit het onderzoek blijkt dat slechts 20,5% vertalers machine vertalen als iets positiefs beschouwen, verder 42,9% als iets negatiefs.

Figuur 36 Machine Translation: opinie van de vertalers



De enquête vraag luidde: *Gebruikt u programma’s voor Machine Vertalen (MT) samen met CAT tools?* Programma’s voor machine vertalen worden vaak in de CAT tool geïntegreerd. Worden ze echter door vertalers gebruikt? Uit de bovenstaande grafiek 36 blijkt dat de meeste vertalers machinevertalen niet gebruiken, namelijk 86%. Slechts 14% vertalers hebben toegegeven dat ze machinevertalen gebruiken namelijk twee vertalers NL-EN / EN-NL en één met beide combinaties. Er zijn enkele opinies:

Enkele opinies van de vertalers met twee kleine:

Ik vertrouw meer op mijn eigen ervaring en kennis. (enquête nr 13)

Poor quality. (enquête nr 1)

Enkele opinies van de vertalers met een kleinere en een grotere taal:

Ik gebruik ze absoluut niet, ze werken beperkend en verlammen je brein. (enquête nr 15) ...
More trouble editing after mt than translating it myself. (enquête nr 2) ... *Nog niet voldoende ontwikkeld/betrouwbaar.* (enquête nr 5)

Enkele positieve reacties:

Google translate om ideeën op te doen /synoniemen te zoeken. (enquête nr 4)

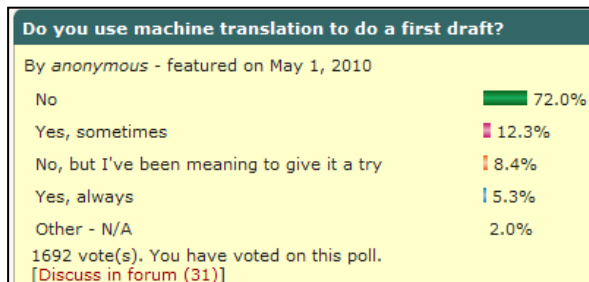
Google translate is integrated in memoq. I also use bing, via an autohotkey script. / because in my language pair google translate often has some good suggestions. (enquête nr 11)

Enkele opinies van de vertalers met beide combinaties:

Google translate, maar alleen tussen Tsjechisch en Slowaaks, of anders ter inspiratie voor de woordenschat. Is ietsje sneller dan losse woorden opzoeken (wat sowieso noodzakelijk blijft, vooral met google). (enquête nr 8) ... *Dat lijkt me weinig nuttig.* (enquête nr 21)

De opinies van de vertalers zijn vooral negatief. Hoewel programma's voor machine vertalen vaak in hun CAT tool geïntegreerd zijn, vertrouwen ze machinevertalen meestal niet en geven ze voorkeur aan eigen kennis. Het kan echter als een inspiratie dienen om ideeën op te doen.

Figuur 37 *Do you use machine translation to do a first draft? (ProZ)*



De resultaten kunnen met een enquête van ProZ.com vergeleken worden. De enquête vraag was: *Do you use machine translation to do a first draft?* De resultaten zijn gelijksoortig. Meerdere vertalers gebruiken machinevertalen niet, namelijk 72% gebruikt het niet en 8,4% wil het graag proberen, verder is er 12, 3% vertalers die machine vertalen soms gebruiken en slechts 5,3% gebruikt het altijd.

6.9.1 Zullen vertalers slechts “post-editors” worden?

Figuur 38 *Will machine translation ever replace human translators? (ProZ)*



Er werd een enquête verricht die genoemd werd: *Will machine translation ever replace human translation?* De resultaten zijn volgende – 69,3% vertalers denkt dat niet, verder 19,7% weet niet en slechts 7,7% denkt dat wel.

De laatste enquête vraag luidde: *Machine Vertalen verbetert zich elke dag – wat is dan de toekomst van vertalers? Zullen ze in de toekomst ‘slechts’ post-editors worden?*

Enkele opinies van de vertalers met twee kleine talen:

De toekomst zal het laten zien. (enquête nr 17) ... Niet zo snel. (enquête nr 1) ... Neen, de vertaler zal nooit worden vervangen, omdat een taal leeft. (enquête nr 13).

Zelf geloof ik, dat een echte vertaler niet kan verdwijnen, hij-zij zal altijd meer waarde betekenen. (enquête nr 17)

Enkele opinies van de vertalers met een kleinere en een grotere taal:

De taak van de vertalers verandert zeker met de opkomst van mt, maar naar mijn mening blijven ze even belangrijk. Het takenpakket verschuift inderdaad meer in de richting van post-editing, maar dit is ook onmisbaar. Bovendien zullen er altijd teksten zijn die bij voorkeur manueel vertaald worden door het belang van accuraatheid (vb. wetteksten voor de eu). (enquête nr 6)

... Het is heel goed mogelijk, zelfs bijzonder waarschijnlijk dat dergelijke programma's ooit vertalers overbodig zullen maken. Als ik de snelheid van de evolutie van de afgelopen 30 jaar bekijk, dan denk ik dat we veilig kunnen stellen dat we dit zelf niet meer zullen meemaken. (enquête nr 18)

Ik ben bang dat mt steeds meer gebruikt gaat worden door bedrijven die weinig geld over hebben voor een goede vertaling, of geen behoefte hebben aan een goede vertaling. Mogelijk ontstaan er twee markten: één voor mt (handleidingen, vluchtige facebook-games etc.) en één voor kwalitatief goede 'echte' vertalingen. (enquête nr 15)

Tot nu toe zijn mt's nog steeds niet goed genoeg om vertalers te vervangen, dus ik denk dat het nog een hele tijd gaat duren voor we alleen post-editors worden. (enquête nr 9)

Enkele opinies van de vertalers met beide combinaties:

Ik gebruik het niet, dus weet niet, tot welke maat ze effectief zijn in het vertalen. Misschien wel voor de grootste talen, maar kleinere talen zullen nog vrij lang door mensen moeten worden vertaald. hoop ik. (enquête nr 14)

Ik denk het niet. Taal heeft zo veel facetten en nuances, machines kunnen daar nooit volledig grip op krijgen. Bovendien is er ook veel creativiteit voor nodig om een goede en mooie vertaling te maken. Daar heb je een mens voor nodig en dan is vertalen efficiënter dan post-editen, lijkt me. (enquête nr 14)

Taal laat zich moeilijk in regels vangen en verandert snel. De meeste vertalers denken dat vertalers altijd nodig zullen zijn omdat enkele soorten vooral belangrijke en officiële teksten moeilijk door machine's vertaald kunnen worden. Dat zijn echter opinies en verwachtingen van vertalers en realiteit kan in werkelijkheid wat anders zijn. Het is noodzakelijk om te beseffen dat steeds meer documenten vertaald moeten worden en als er geen goede kwaliteit nodig is, maakt het machinevertalen namelijk Google translate mogelijk. TAUS zegt over Google Translate:

Google's computers ... produce ten times more translated words than the entire professional translation workforce worldwide.¹²²

Verder volgens een voorstel voor een verordening van de Raad tot het aangaan van nauwere samenwerking op het gebied van de instelling van eenheidsoctrooibescherming met betrekking tot de toepasselijke vertaalregelingen:

Gezien de technologische ontwikkeling mag de maximumperiode voor de ontwikkeling van hoogwaardige machinevertalingen niet worden geacht langer dan 12 jaar te duren.¹²³

De toekomst zal laten zien of en hoe de taken van vertalers zullen veranderen. Het zou zeker interessant zijn om het ontwikkelen van machinevertalen te volgen en verder te onderzoeken.

¹²²Jaap Van Der Meer. 2 august 2010. "Where are Facebook, Google, IBM and Microsoft taking us?" TAUS. 22 april 2012 <<http://www.translationautomation.com/perspectives/where-are-facebook-google-ibm-and-microsoft-taking-us.html>>.

¹²³ 2011/0094 (CNS). 13.4.2011, Brussel. p.12. 30 mei 2012
<<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0216:FIN:nl:PDF>>.

7 CONCLUSIE

Het hoofdthema van de scriptie is computerondersteund vertalen. De nood aan vertalingen is vooral dankzij globalisering en het internet voortdurend aan het stijgen. Ontwikkelingen zijn niet stop te zetten en daarom is het voor een vertaler noodzakelijk om op de hoogte te zijn van beschikbare vertaalprogramma's en nieuwe taaltechnologieën. Uit het onderzoek blijkt dat steeds meer vertaalbureau's van hun vertalers verwachten dat ze vertaalprogramma's kennen en kunnen gebruiken. Zonder kennis van zulke programma's wordt het vooral voor nieuwe vertalers lastig om werk te krijgen. Taaltechnologieën die ontwikkeld worden, betekenen voor vertalers een andere manier van werken. Vertaalgeheugens en glossaria maken het werk consistent, vlugger en efficiënter. Verder zijn er programma's voor machinevertalen die zelf hele teksten vertalen en optische tekenherkenning die het mogelijk maakt om uitgeprinte teksten in een computer weer te kunnen bewerken. Het is echter niet zwart-wit. Zoals het vaak het geval is, hebben ook taaltechnologieën naast voordelen ook enkele nadelen, de opinies van vertalers op dit gebied verschillen bovendien vaak van elkaar.

De scriptie bevat twee belangrijke onderdelen: een theoretisch en een praktisch deel. In het theoretische deel worden belangrijke termen uitgelegd en ernaast ook verschillende taaltechnologieën behandeld. Het is belangrijk om het verschil tussen computerondersteund vertalen en machinevertalen te beseffen, verder ook het verschil tussen vertalen en lokalisatie. Deze termen worden vaak door elkaar gebruikt (soms door vertalers zelf) en het leidt tot misverstanden en wantrouw tegen programma's. Zoals het al uit het woord zelf duidelijk is, "steunt" computerondersteund vertalen eigenlijk alleen een vertaler, hij/zij moet het werk verder zelf doen, glossaria en vertaalgeheugens spelen hier verder een belangrijke rol.

In hoofdstuk 4 worden de meest gebruikte vertaalprogramma's besproken namelijk Trados, Wordfast, MemoQ en Déjà vu. Ze worden grondig besproken en vergeleken op het gebied van hun prijs, aangeboden producten, systeembenodigdheden, ondersteunde talen en bestandsformaten. Het volgende begrip is lokalisatie als een soort "modern" vertalen. Meestal gaat het om vertalen van webpagina's. Lokalisatie betekent niet alleen vertalen maar ook een product aan een doelmarkt aan te passen en namelijk geschikte maten, datumformaten enz. te gebruiken maar ook verschillend technisch vermogen om met programma's te werken en formaat van de tekst te bewaren, zijn noodzakelijk. Verder zijn er programma's voor machine vertalen die zelfstandig werken, een vertaler is hier eigenlijk niet meer nodig, hij/zij kan als een "post-editor" dienen. Op dit moment zijn deze programma's nog steeds gebrekkig. Ze worden alleen gebruikt voor enkele soorten teksten namelijk repetitieve teksten met vaste terminologie bijvoorbeeld weerberichten. Op dit moment kunnen ze vertalers niet vervangen, ze kunnen hoewel met vertalers goed complementair werken. Een taal is een dynamisch systeem die in de regels niet te bevatten is. Talen verschillen zich van elkaar en context en wereldkennis zijn noodzakelijk om een taal volledig te begrijpen. Om deze reden is het voor

computers onmogelijk om een taal volledig te beheersen. Computers ontbreken pragmatische kennis die aan mensen helpt om ambiguïteit te vermijden of op te lossen. Verder is er ook een nieuw begrip *crowdsourcing* wat een soort “gemeenschappelijk” vertalen is. We komen dit fenomeen steeds vaker tegen (vaak onbewust) bijvoorbeeld op Wikipedia, Facebook of Twitter. Leden of paginagebruikers, meestal geen professionele vertalers, worden gevraagd om een vertaling het liefst naar de moedertaal te verrichten. Vrijwilligers verlenen zo gratis vertalingen aan deze pagina’s. De vertalingen zijn niet perfect wat geen probleem is omdat andere leden fouten gewoon gaan verbeteren. Belangrijk is dat alles gratis, op vrijwilligersbasis werkt. Dat kan samen met machinevertalen een soort concurrentie voor vertalers betekenen omdat ze op zo’n manier betaald werk verliezen.

Het belangrijkste deel van de scriptie was een onderzoek, namelijk een enquête die voor vertalers die met CAT tools werken samengesteld werd. 21 vertalers hebben aan de enquête deelgenomen, en wel 10 vertalers met de talencombinatie NL-CS, CS-NL en 11 vertalers EN-NL, NL-EN. De centrale vraag van de scriptie luidt: *Wat voor verschillen zijn er tussen computerondersteund vertalen van twee kleinere talen (NL-CS, CS-NL) en één kleinere en één grote taal (EN-NL, NL-EN)?* Vertalers werden ondervraagd welke programma’s ze gebruiken en waarom. Welke opleiding ze hebben afgerond en of ze in een bepaald soort teksten gespecialiseerd zijn. Ook hun opinie over machinevertaling en andere vragen waren van belang. Om de resultaten te ondersteunen en vergelijken werden ook andere onderzoeken gebruikt onder vertalers die met CAT tools werken, vooral korte enquêtes beschikbaar op www.ProZ.com. Er zijn enkele interessante feiten en verschillen tussen talencombinaties gebleken, namelijk:

- Een groot aantal vertalers die tussen twee kleinere talen vertalen heeft nog andere taalcombinaties, in dit geval is het 50% van de vertalers. Op ProZ.com heeft 71% vertalers met combinatie NL-CS, CS-NL meerdere taalcombinaties hebben met een grotere taal die vooral Engels is.
- Het meest gebruikte programma is SDL Trados (36 %) hoewel dit het duurste programma is. De volgende programma’s zijn MemoQ (15%), Wordfast en WordfastPro beide met 12%. Verder Déjà vu en Across met 7%, Strar Transit en Idiom met 5% en SDLX Edit met 3%.
- Vertalers tussen geven EN/NL de voorkeur aan Trados terwijl vertalers NL/CS een voorkeur aan Wordfast geven. De reden kan zijn dat vertalers NL/CS niet zo veel geld aan het programma’s willen uitgeven de redenen zijn dat ze grotendeels één programma hebben en dat is vaak Wordfast, dat aanzienlijk goedkoper is dan Trados.
- Drie belangrijkste redenen om een CAT tool te kopen zijn de volgende: omdat het door een vertaalbureau vereist wordt (30%); dat het programma gebruiksvriendelijk is (17%) en tenslotte op de derde plaats (13%) een overzichtelijke lay-out.

- Er is een tendentie dat vertalers tussen kleinere talen zich het werk zonder CAT tools beter kunnen voorstellen. De opinie dat vertalen zonder CAT tools niet meer kan, heeft in totaal echter de overhand met 59% stemmen.
- Uit het onderzoek blijkt dat vertalers actief op het internet CAT tools vaker gebruiken (78,2% op ProZ.com). Verder geeft een groot vertalers uit Centraal Europa, vooral uit Tsjechië nog steeds vaak de voorkeur aan werken zonder vertaalprogramma's (51% volgens JTP).
- De belangrijkste voordelen van CAT zijn volgende: vertaalgeheugen, tijdbesparing, vaste terminologie en hogere productiviteit.
- Het grootste nadeel is dat CAT tools niet voor alle soorten teksten geschikt zijn (25%). Verder zijn dat technische problemen (21%), prijs van het programma (16%), het gebruik van het programma is niet intuïtief (12%) en weinig creatieve taal 9%.
- 57% vertalers zeggen dat het vertaalgeheugen belangrijker is, 24% glossarium en 19% zeggen dat beide even belangrijk zijn. Het is duidelijk dat combinatie NL/CS meer waarde hecht aan glossarium dan NL/EN. De verklaring is als volgt: voor de combinatie NL/CS bestaat er echter niet voldoende woordenboek en gespecialiseerde woordenboeken zijn er helemaal niet. Voor een vertaler is het daarom lastiger om terminologie van goede kwaliteit aan te schaffen.
- Noodzakelijke competenties van een vertaler zijn natuurlijk een perfecte kennis van doel- en brontaal en beide culturen. Verder zijn dat computervaardigheden, algemene kennis, brede interesse, specialisatie, communicatieve vaardigheden en onderhandelen. Een vertaler moet bereid zijn om voortdurend te leren en ontwikkelen. Wat eigenschappen betreft zijn zeker perfectionisme, geduld, flexibiliteit, betrouwbaarheid en klantvriendelijkheid van belang.
- Het is opvallend dat een groot aantal vertalers geen taalopleiding heeft (38 %). Het betekent dat 62% van de vertalers een taalopleiding heeft waarvan slechts 33% van de vertalers *Master in het Vertalen of Translatologie* gestudeerd hebben. 90% vertalers NL-CS of/en CS-NL hebben een taalopleiding en bij vertalers NL-EN en/of EN-NL is het slechts 45,5%.
- Uit het onderzoek blijkt dat vertalers met kleinere talen zich vaker bijscholen (vooral extra opleiding of cursussen) dan vertalers met één grotere taal die allemaal een voorkeur aan eigen ervaring en zelfstudie geven. Het is met de opleiding van vertalers verbonden.
- De meest voorkomende typen te vertalen teksten zijn economische, handleidingen, juridische en technische.
- Met lokalisatie is 73% van de vertalers tussen één grotere en één kleinere taal bezig maar slechts 30% van de vertalers tussen twee kleine talen en vertalers die beide combinaties hebben. Engels is een belangrijke taal en steeds meer materialen worden naar/vanuit het

Engels vertaald daarom is de vraag hoog, dat wil zeggen aanzienlijk hoger dan tussen twee kleinere talen.

- Het blijkt dat de meeste vertalers geen machinevertalen gebruiken namelijk 86%. De opinies van de vertalers zijn vooral negatief. Hoewel programma's voor machinevertalen vaak in hun CAT tool geïntegreerd zijn, vertrouwen ze machinevertalen meestal niet en geven ze de voorkeur aan eigen kennis. Machinevertalen kan echter als een inspiratie dienen om ideeën op te doen.
- De meeste vertalers denken dat vertalers altijd nodig zullen zijn omdat enkele soorten vooral belangrijke en officiële teksten moeilijk door machine's vertaald kunnen worden.

Het is noodzakelijk om te beseffen dat steeds meer documenten vertaald moeten worden en als er geen hoge kwaliteit verwacht wordt, doet machinevertalen het werk goed genoeg. De toekomst zal het echter laten zien of en hoe de taken van vertalers zullen veranderen. Het zou zeker interessant zijn om het ontwikkelen van vertaaltechnologieën en machinevertalen verder te volgen en onderzoeken.

Resumé

Hlavním tématem diplomové práce je počítačem podporovaný překlad s důrazem na překladatele z nizozemštiny do češtiny a angličtiny. Práce se skládá ze dvou částí – teoretické a praktické. V teoretické části je čtenář seznámen s nejdůležitějšími pojmy a vývojem počítačem podporovaného překladu, dále jsou zde představeny a porovnány nejpoužívanější programy (Trados, Wordfast, Memoq a *Déjàvu*). V práci jsou dále zmíněny další pojmy a technologie nezbytné pro překladatele – strojový překlad (Google Translate a Babelfish), lokalizace, crowdsourcing a optické rozpoznávání znaků (OCR).

Praktická část se skládá z ankety určené překladatelům mezi dvěma menšími jazyky (NL-CS, CS-NL) a jedním menším a jedním větším jazykem (NL-EN, EN-NL). Hlavní důraz je kladen na názory překladatelů a na případné rozdíly mezi oběma skupinami, jako například jaké programy používají, výhody a nevýhody daných programů, jejich nezbytnost a využití. Dále vzdělání a specializace překladatelů, jejich názor na strojový překlad, lokalizaci a další. Výsledky ankety jsou porovnány a podloženy dalšími výzkumy a anketami zaměřenými na počítačem podporovaný překlad.

Summary

The main topic of the final work is Computer Aided Translation focused mainly on translators who translate from Dutch to Czech and/or English. The work consists of two parts – a theoretical and a practical one. In the theoretical part, the most important terms are introduced, as well as the development and history of Computer Aided Translation, and finally the most used programmes are presented and compared (Trados, Wordfast, Memoq and Déjà vu). Subsequently, other important terms and technologies necessary for translators are introduced – Machine Translation (Google Translate, Babelfish), Localisation, Crowdsourcing and Optical Character Recognition.

The practical part is based on questionnaires for translators between two smaller languages (NL-CS, CS-NL) and one bigger and one smaller language (EN-NL, NL-EN). An emphasis is placed on the opinions and experience of translators and potential differences between both groups. They were asked which programmes they use, what are the advantages and disadvantages of each, and their usefulness and indispensability. Further concerns are the education of translators and their specialisation, their view on machine translation and localisation, and more. The results are compared and supported by other questionnaires and research dealing with Computer Aided Translation.

ANOTACE

Jméno: Lucie Kadeřábková

Název katedry a fakulty: Katedra nederlandistiky FF UP v Olomouci

Název diplomové práce: Computerondersteundvertalentussen twee kleineretalen (NL-CS, CS-NL) en éénkleinere en één groteretaal (EN-NL, NL-EN)

Vedoucí diplomové práce: doc. dr. Wilken Engelbrecht, cand. litt.

Počet znaků (bez mezer): 135 961

Počet stran: 112

Počet příloh: 2

Klíčová slova: computerondersteund vertalen, CAT tools, vertaaltechnologieën, machine vertalen, enquête voor vertalers

Krátká charakteristika:

Hlavním tématem diplomové práce je počítačem podporovaný překlad. Práce se skládá ze dvou částí – teoretické a praktické. V teoretické části je čtenář seznámen s nejdůležitějšími pojmy, vývojem strojového a počítačem podporovaného překladu, dále jsou zde představeny a porovnány nejpoužívanější programy (Trados, Wordfast, Memoq a *Déjàvu*). Praktická část se skládá z ankety určené překladatelům mezi dvěma menšími jazyky (NL-CS, CS-NL) a jedním menším a jedním větším jazykem (NL-EN, EN-NL). Hlavní důraz je kladen na názory překladatelů a na případné rozdíly mezi oběma skupinami, jako například jaké programy používají, výhody a nevýhody daných programů, dále vzdělání a specializace překladatelů, jejich názor na strojový překlad a další.

Bronnen

- ADAMS, Douglas. *The Hitchhiker's Guide to the Galaxy*. NY: Harmony books, 1979.
- ADRIAENS, Geert, Steven et al. *Lexical ambiguity resolution: Perspectives from Psycholinguistics, Neuropsychology, and Artificial Intelligence*. CA: Morgan Kaufman Publishers, 1988.
- BART, Peter van, M. Steehouder. *Basisboek: Technische Communicatie*. Assen: Koninklijke Van Gorcum, 2008.
- BECKER, H.A. *Generaties en de strijd tegen taalbarrières*. Utrecht: Universiteit Utrecht, 2001.
- DAWOENS, Gudrun. *Taal aan den lijve: Het gebruik van corpora in taalkundig onderzoek en taalonderwijs*. Gent: Academia Press: 2008.
- DOUGLAS, Arnold, Lorna Balkan et al. *Machine Translation: An Introductory Guide*. London: NCC Blackwell, 1994.
- GRADDOL, David. *Future of English: A guide to forecasting the popularity of the English language in the 21st century*. The British Council, 2000.
- GOMÉZ, Pascual Cantos . *Lexical ambiguity, dictionaries and corpora*. Universidad de Murcia, 1996.
- HOOGENDOORN, Hans, Irene van Vuuren-Matejiwkova *ANWB taalgids Tsjechisch*. ANWB Media boeken, 2004.
- MIKULÍK, Petr. *A Study of selected Machine Translation Systems*. Diplomová práce. Brno: Masaryk Universiteit, 2009.
- POPELA, Jaroslav, Vykypěl Bohumil, Boček Vít. *Skaličkova jazyková typologie*. Brno: Masarykova univerzita, 2006.
- QUAH, Chiew K. *Translation and Technology*. Basingstoke: Palgrave Macmillan, 2006.
- SHIBATANI, Masayoshi, Theodora Bynon. *Approaches to language typology*. Oxford: Oxford University Press, 1995.
- STERKENBURG, P. G. J. van. *Unity and diversity of languages* . Amsterdam: John Benjamins Publishing Co. 2008.
- THIJSSSENS, Jos, Eric Tytgat, Bart Vandenberghe. *Taalkracht tien 4 – bronnenboek*. Malle: De Sikkel, 1996.
- VANDEWEGHE, Willy. *Duoteksten: Inleiding tot vertaling en vertaalstudie*. Gent: UGent, 2005.

Elektronische bronnen

- ARNOLD et al. 1994. *Machine Translation: An Introductory Guide*. 28 maart 2012 <<http://www.ling.ohio-state.edu/~dickinso/384/wi04/slides/mt-slides-4up.pdf>>.
- “Automated Translation.” SDL. 8 april 2012 <<http://www.translationzone.com/en/translator-solutions/automated-translation/>>.
- BENITO, Gloria Álvarez. 1994. *Contextual Deletion of Object and Ambiguity in Machine Translation*. 25 maart 2012 <http://rua.ua.es/dspace/bitstream/10045/6012/1/RAEI_07_02.pdf>.
- BOUMA, Gosse, Gertjan van Noord. 1998. “Natuurlijke-taalinterfaces“. Vakgroep Alfa-informatica & BCN Rijksuniversiteit Groningen. 14 februari 2012 <<http://www.let.rug.nl/vannoord/papers/breinmakers/nli.html>>.
- ESSENLINK, Bert. 2012. “Evolution of Localisation“. 18 februari 2012 <<http://www.scribd.com/doc/12596514/The-Evolution-of-Localization>>.
- HUTCHINS, John. 1991. “Why computers do not translate better.” 5-6. 5 april 2012 <<http://www.mt-archive.info/Aslib-1991-Hutchins.pdf>>
- JEBBAR, Abdelhak 2012. *Machine Translation vs human Translation*. 5 april 2012 <<http://www.translationdirectory.com/articles/article1326.php>>.
- KUIPER, Ineke. 2008. “Vertaalprogramma’s“. 18 februari 2012 <<http://translations.inekekuiper.nl/vertaalprogramma%27s.htm>>.
- LEWIS, M. Paul. 2009. *Ethnologue: Languages of the World, Sixteenth edition*. 13 februari 2012 <http://www.ethnologue.com/ethno_docs/distribution.asp?by=size>.
- “MemoQ Server“. 2012. Kilgray Translation Technologies. 28 maart 2012 <<http://kilgray.com/products/memoq-server>>.
- NIJHOLT, A. 1994. “Parlevink in kontekst: meer dan language engineering“. 25 maart 2012 <<http://doc.utwente.nl/67128/1/eniac.pdf>>.
- O’HAGAN, Minako. “Training for localisation“. School of Applied Language and Intercultural Studies Dublin City University, Ireland. 5 februari 2012 <http://isg.urv.es/library/papers/OHagan_Training.pdf>.
- PRIOR, Marc. 2011. “Optical karakter recognition“. 9 februari 2012 <<http://www.marcprior.de/linux/misc/ocr.html>>.

- RENCKENS, Erica. 23 februari 2009. "I sit with the hands in the hair". Kennislink. 8 februari 2012 <<http://www.kennislink.nl/publicaties/i-sit-with-the-hands-in-the-hair>>
- "Een terminologielijst samenstellen". November 2011. Tilburg translation. 18 februari 2012 <<http://www.tilburgtranslations.nl/tip-van-de-maand/>>.
- "Google Translation". 2009. Vertaalwerk. 2 februari 2012 <<http://vertaalwerk.jouwpagina.nl/>>.
- "Google Translate: Gratis online vertalen met Google". 1.4. 2010. PC and Internet: Tips and Tricks. 22 februari 2012 <<http://pc-en-internet.infonu.nl/tips-en-tricks/53497-google-translate-gratis-online-vertalen-met-google.html>>.
- "Homonymie, polysemie en synonymie." 30 oktober 2011. Kunst en cultuur 25 maart 2012 <<http://kunst-en-cultuur.infonu.nl/taal/85815-homonymie-polysemie-en-synonymie.html>>.
- "How Google Translate Works. 13 september 2011. The independent. 18 maart 2012 <<http://www.independent.co.uk/life-style/gadgets-and-tech/features/how-google-translate-works-2353594.html>>.
- "OCR". 2012. CVISION Technologies. 9 februari 2012 <<http://www.cvisiontech.com/reference/general-information/ocr.html?lang=dut>>.
- "Překladač Google". 2011. Google. 20 februari 2012 <http://www.google.nl/help/faq_translation.html>.
- "SDL MultiTerm Extract". SDL. 8 april 2012 <<http://www.translationzone.com/en/translator-products/sdlmultitermextract/>>.
- "SDL MultiTerm Desktop". SDL. 8 april 2012 <<http://www.translationzone.com/en/translator-products/sdlmultitermdesktop/>>.
- "SDL Passolo 2011". SDL. 8 april 2012 <<http://www.translationzone.com/en/translator-products/sdl-passolo/>>.
- "Stand-alone versus embedded CAT tools: trade-offs" 6 november 2008. Felix blog. 19 maart 2012 <<http://felix-cat.com/blog/2008/11/06/stand-alone-versus-embedded-cat-tools-trade-offs/>>.
- "Translator Group buying". 2012. ProZ. 9 april 2012 <http://www.proz.com/?sp=store&sp_mode=overview>.

- *Verbuiging*. 2012. Nederlandse Taalunie, Taaladvies.net. 9 februari 2012
<http://woordenlijst.org/leidraad/lijst_van_vaktermen/verbuiging/>.
- “Vertaalgeheugensprogramma’s“. Vertalers Platform. 18 februari 2012<<http://www.vertalersplatform.nl/iv-vertaalgeheugenprogs.php>>
- “Vertalen met Babel Fish: De online-tolk bij uitstek“. 6 juli 2006. Clickx. 19 februari 2012 <<http://www.clickx.nl/hintsentips/58376/vertalen-met-babel-fish/>>.
- “Vertaalgeheugens“. 2011. Industrial services. 5 februari 2012
<http://www.be.sgs.com/nl_be/translation-memories.htm?selen=1&serviceid=10143648&lobid=5550>.
- “Very Large Translation Memory Project“. 2012. Wordfast. 27 maart 2012
<http://www.wordfast.com/products_vltm.html>.
- *Wat is ProZ.com?* 2012.ProZ. 30 maart 2012
<http://dut.proz.com/faq/general.html#what_is_proz_com>.
- “What is the difference between translation and localisation?“ 2009. Translation services. 11 februari 2012 <<http://www.translation-service.com/faqs/localization/>>.
- “WinAlign“. 8 september 2007. Taggers help. 11 februari 2012
<<http://taggershelp.blogspot.com/2007/09/winalign.html>>.
- “Wordfast Pro – The next step in translation Memory technology. Wordfast. 25 maart 2012“<http://www.wordfast.com/docs/WF_ProProductbrief.pdf>.
- “Wordfast Classic“. 2012. Wordfast. 27 maart 2012
<http://www.wordfast.com/products_wordfast.html>.
- “Wordfast Anywhere“. 2012. Wordfast. 27 maart 2012
<<http://www.wordfast.net/?whichpage=anywhere>>.

BIJLAGEN

Bijlage I *Enquête voor vertalers NL-CS, CS-NL of NL-EN, EN-NL die CAT-tools gebruiken*

Bijlage II *Enquêtes beschikbaar op <<http://www.proz.com/polls>>.*

BIJLAGE I

Vraagstelling: *Wat voor verschillen zijn er tussen computerondersteund vertalen van twee kleinere talen en één kleinere en één grote taal?*

Olomouc, 17 december 2011

Geachte heer/mevrouw,

Mijn naam is Lucie Kadeřábková. Ik studeer Nederlandse en Engelse Filologie aan de Palacký Universiteit in Olomouc, Tsjechië. Ik schrijf mijn masterscriptie over *computerondersteund vertalen*. Een belangrijk onderdeel van mijn onderzoek is gebaseerd op deze enquête.

Doel van de scriptie

De scriptie zal een theoretisch en praktisch gedeelte bevatten. In het theoretische gedeelte zal ik de problematiek van *computerondersteund vertalen* bespreken. In het praktische gedeelte wil ik graag verschillende programma's vergelijken op het gebied van de prijs, overzichtelijkheid, aangeboden producten, ondersteuning van de talen en bestandsformaten. Het belangrijkste onderdeel zal echter de analyse op de grond van deze enquête zijn.

Doel van deze enquête

Deze vragenlijst is bedoeld voor vertalers die met CAT-tools werken en dan vooral in de talencombinatie(s) NL-CS, CS-NL of NL-EN, EN-NL. Ik wil graag onderzoeken of er verschillen zijn tussen computerondersteund vertalen tussen twee kleinere talen en één kleinere en één grotere taal vooral in verband met de keuze van het programma, de opleiding, vakspecialisatie of de te vertalen teksten.

Verdere informatie

Het welslagen van het onderzoek hangt af van uw medewerking. Er wordt slechts **10 minuten** van uw tijd gevraagd. Het zou mij erg helpen als u de ingevulde enquête **vóór 31 januari 2012** terug zou willen sturen naar lucie.ltm@gmail.com. Deze enquête is anoniem. U kunt deze vragenlijst dus in alle vertrouwen invullen.

Bij voorbaat dank voor uw hulp. Uw bijdrage vormt de basis voor het succesvol afronden van mijn onderzoeksproject. Geïnteresseerden in de resultaten kunnen mij via email lucie.ltm@gmail.com contacteren.

Bedankt voor uw medewerking!

Met vriendelijke groeten,

Lucie Kadeřábková

- 1) **Tussen welke talen vertaalt u?**
- | | | |
|----------|----------|----------|
| a) NL-CS | c) NL-EN | e) NL-SK |
| b) CS-NL | d) EN-NL | f) SK-NL |
- Andere talencombinatie(s).....
- 2) **Welke CAT-tool(s) gebruikt u en hoe lang heeft u hiermee ervaring?**
- | | | | |
|-----------------|------------|-------------------|------------|
| a) SDL Trados | jaar | e) Star Transit | jaar |
| b) WordFast | jaar | f) Omega T | jaar |
| c) WordFast Pro | jaar | g) Memoq | jaar |
| d) Déjà Vu | jaar | h) Andere(n)..... | |
- 3) **Om welke reden(en) heeft u bepaalde CAT-tool(s) gekozen? Kies maximaal 4 opties.**
- | | |
|-----------------------------------|----------------------------------|
| a) Door een vertaalbureau geëist | d) Overzichtelijke lay-out |
| b) Door een klant geëist | e) Tool is gebruiksvriendelijk |
| c) Lage prijs | f) Brede ondersteuning van talen |
| g) Verschillende bestandsformaten | h) Andere(n) |
- 4) **Wat zijn de voordelen van CAT-tools (uw eigen ervaring)? Kies maximaal 4 opties.**
- | | |
|------------------------------------|---|
| a) Tijdsbesparing | g) Vereenvoudigen van het werk |
| b) Hogere productiviteit | h) Ondersteuning van verschillende bestandsformaten |
| c) Coherentie van de teksten | i) Gebruik van het programma intuïtief |
| d) Vaste terminologie (glossarium) | j) Andere(n)..... |
| e) Vertaalgeheugen (TM) | |
| f) Prijs van het programma | |
- 5) **Wat zijn de nadelen van CAT-tools (uw eigen ervaring)? Kies maximaal 4 opties.**
- | | |
|------------------------------------|-------------------------------|
| a) Niet voor alle soorten teksten | f) Technische problemen |
| b) Prijs van het programma | g) Gebruik niet intuïtief |
| c) Vereenvoudigen van de taal | h) Onbrekende functie(s)..... |
| d) Weinig creatieve taal | i) Andere(n)..... |
| e) Technisch vermogen noodzakelijk | |
- 6) **Is volgens u vertalen voor de zakelijke markt nog mogelijk zonder CAT-tools?**
- a) Ja
- b) Nee
- Waarom?.....
-
- 7) **Wat vindt u belangrijker / Wat gebruikt u vaker – glossarium of vertaalgeheugen?**
- Waarom?**
-
-
- | | |
|---|--|
| • Hoe krijgt u meestal een glossarium (Kies maximaal 3 opties) | • Hoe krijgt u meestal een vertaalgeheugen (Kies maximaal 3 opties) |
| a) Klant | a) Klant |
| b) Vertaalbureau | b) Vertaalbureau |
| c) Andere vertaler | c) Andere vertaler |
| d) Kopen | d) Kopen |
| e) Gratis – via het internet | h) Gratis – via het internet |
| f) Zelf maken | e) Zelf maken |
| g) Anders..... | f) Anders..... |

8) **Welke competenties zijn noodzakelijk om tegenwoordig vertaler te worden?**
(behalve talenkennis).....
.....

9) **Welke opleiding heeft u afgerond?**
a) Master in het Vertalen e) Geen
b) Taal- en Letterkunde f) Andere.....
c) Translatologie
d) Filologie

10) **Welke soort(en) teksten vertaalt u?**
a) Technische e) Handleidingen
b) Juridische f) EU
c) Economische g) Andere (n).....
d) Medische

11) **Heeft u zich in een bepaald soort teksten gespecialiseerd?**
a) Ja
b) Nee
• Zo ja, hoe? - Extra opleiding / Cursussen / Eigen ervaringen / Anders

12) **Is het voor vertalers mogelijk zich in een bepaald soort teksten te specialiseren? Zijn er hiervoor wel genoeg teksten?** (vooral bij kleinere talen)?
a) Ja
b) Nee
• Waarom?.....
.....

13) **Als u vaak verschillende soorten teksten vertaalt, zijn CAT-tools dan voor u nuttig?**
a) Ja
b) Nee
• Waarom?.....
.....

14) **Houdt u zich ook bezig met vertalen van webpagina's (lokalisatie)?**
a) Ja
b) Nee
• Zo ja, welk(e) programma(s) gebruikt u?.....

15) **Gebruikt u programma's voor Machine Vertalen (MT) samen met CAT-tools?**
c) Ja
d) Nee
• Zo ja, welk en waarom? / Zo niet, waarom?

16) **Machine Vertalen verbetert zich elke dag – wat is dan de toekomst van vertalers? Zullen ze in de toekomst “slechts” post-editors worden? Uw opinie, aub:**
.....
.....
.....

Bedankt voor uw tijd!
Lucie Kadeřábková

BIJLAGE II – Enquêtes beschikbaar op <<http://www.proz.com/polls>>.

