

# Publiceren en vertalen

## Problemen en oplossingen

Menno van Zaanen  
studentnummer: 9497188  
1998



UNIVERSITEIT VAN AMSTERDAM

**ALFA****BASE**  
PUBLIKATIE PROCESSORS



De antwoorden zijn altijd al aanwezig  
—Steve Vai



# Abstract

Deze scriptie beschrijft de resultaten van een onderzoek naar de problemen en oplossingen van de combinatie van een publicatie- en vertaaltraject.

## Probleemstelling

Het publicatie- en vertaaltraject bestaan beide reeds jaren. Bij de combinatie van de twee trajecten wordt nu meestal nog zeer handmatig gewerkt. De vraag is of deze trajecten beter op elkaar aangesloten kunnen worden. Ook is van belang waarom dit mogelijk of niet mogelijk is en als het mogelijk is hoe dit dan moet gebeuren. Daarnaast is onderzoek gedaan naar de mogelijke hulpmiddelen van de trajecten.

Om goed op de toekomst voorbereid te zijn is ook een onderzoek gestart naar de mogelijke veranderingen in de toekomst. Met deze kennis is beter in te springen op de toekomstige veranderingen en mogelijke problemen kunnen met deze kennis nu of in de toekomst opgelost worden.

## Conclusie

Bij de combinatie van het publicatie- en vertaaltraject kan een bedrijf als Alfa▼Base een spilfunctie krijgen. Alfa▼Base verzorgt de nodige documentlogistiek en daarnaast het publicatietraject.

## Uitbreiding

Het publicatietraject kan, zoals het nu is, gemakkelijk uitgebreid worden met een vertaalstap, vooral wanneer een bedrijf wordt ingeschakeld dat vertalingen naar meerdere talen kan afhandelen. Een dergelijk bedrijf organiseert de vertalingen en kan de logistiek van de vertalingen afhandelen, terwijl Alfa▼Base de rest van het gehele traject afhandelt.

## Documentlogistiek

Wanneer documenten gepubliceerd moeten worden in meerdere talen, ontstaan veel verschillende versies van de documenten. Een goed versiebeheer is essentieel om de explosieve groei van documenten in bedwang te houden. Dit is een van de grootste problemen bij het combineren van de beide trajecten.

Er moet dus programmatuur voor versiebeheer gevonden worden. Dit is mogelijk door onderzoek naar bestaande systemen of door zelf een systeem te ontwerpen.

## **Knelpunten**

Elk project heeft zijn eigen specifieke problemen. Er moet eerst gekozen worden welke hulpmiddelen gebruikt worden. Voor elk soort hulpmiddel moet afgevraagd worden of het hulpmiddel wel gebruikt moet worden, welk programma gebruikt wordt en op welke locatie het hulpmiddel gebruikt moet gaan worden.

## **Database publicatie**

Bij een database publicatie wordt veel, zo niet alle, informatie in een publicatie gehaald uit een database. Dit kan duidelijk gestructureerde informatie zijn zoals bijvoorbeeld technische informatie van bepaalde producten, maar het kan ook delen van tekst zijn, die samengevoegd worden. Vooral de laatste soort database publicatie bestaat nog niet erg lang, maar zal in de toekomst waarschijnlijk veel terrein winnen.

Het voordeel van database publicaties is dat vertalingen direct in de database opgeslagen kunnen worden. Deze vertalingen kunnen samen met de andere informatie uit de database gehaald worden, waarbij vertaalde stukken tekst gecombineerd kunnen worden. Deze stukken tekst hoeven niet opnieuw vertaald te worden.

## **Communicatie**

Hoe goed Alfa▼Base de interne, technische problemen ook heeft opgelost, er blijft altijd het probleem dat meerdere bedrijven tegelijkertijd aan hetzelfde project werken. Om verschillende bedrijven goed met elkaar te kunnen laten samenwerken is een goede communicatie tussen de verschillende bedrijven nodig.

Wanneer Alfa▼Base een vertaalbureau zoekt die vertalingen naar meerdere talen kan produceren, is het mogelijk dit bedrijf voor meerdere opdrachten te gebruiken. Daardoor is het mogelijk een goede relatie met dit bedrijf op te bouwen.

## **Toekomst**

In de toekomst zal er uiteraard wat in het publicatie- en vertaaltraject gaan veranderen. Toch verwachten we voorlopig geen structurele veranderingen. De grootste veranderingen zullen verbeteringen aan bestaande systemen zijn. Eventuele structurele veranderingen in de trajecten zullen geleidelijk gebeuren.

## **Organisatie**

Het vertaaltraject binnen het publicatietraject zal in de toekomst beter georganiseerd worden. Vertaalbureaus hebben dan een beter overzicht over de kosten en tijdsduur van vertalingen. Daarnaast worden vertaler meer en meer geholpen door verschillende vertaalhulpmiddelen.

### **Bestandsformaten**

Door het gebruik van open standaarden voor bestandsformaten wordt de interface tussen het publicatie- en vertaaltraject minder problematisch. Bestanden met vertalingen naar talen met speciale karakters kunnen gemakkelijk uitgewisseld worden waardoor de teksten in de verdere stappen in het publicatietraject zonder problemen verwerkt kunnen worden. Unicode en SGML zijn twee systemen die in de toekomst in dit gebied zeer belangrijk kunnen worden.

### **Vraag en aanbod**

De vraag naar vertalingen zal in de toekomst toenemen. Bedrijven die binnen het publicatietraject een vertaaltraject kunnen integreren zullen dus een voordeel hebben ten opzichte van bedrijven die alleen een publicatietraject kunnen aanbieden. De afzonderlijke organisatie en controle van de twee trajecten kan veel tijd en moeite kosten, dus veel bedrijven zullen hun publicaties laten uitvoeren door een bedrijf dat een combinatie van de twee trajecten kan organiseren.

### **Standaardpakket**

De combinatie van het vertaal- en publicatietraject heeft altijd dezelfde structuur. Hiervoor zou een standaardpakket ontwikkeld kunnen worden. Dit standaardpakket kan gebruikt worden als basis van het publicatie- en vertaaltraject. Dit standaardpakket samen met een goede relatie met een vertaalbureau zorgt voor een goed gestructureerd publicatie- en vertaaltraject. De precieze invulling van alle stappen in het traject kan per project aangepast worden; elk project is immers anders.





# Inhoudsopgave

<b>1</b>	<b>Inleiding</b>	<b>1</b>
<b>2</b>	<b>Huidige situatie</b>	<b>5</b>
2.1	Situaties in de praktijk . . . . .	5
2.2	De koppeling tussen het vertaal- en publicatietraject . . . . .	6
2.2.1	Oplossen van problemen . . . . .	6
2.2.2	Verbeteringen door automatisering . . . . .	8
<b>3</b>	<b>Het publicatietraject</b>	<b>11</b>
3.1	Publiceren . . . . .	11
3.2	Database-publicatietraject . . . . .	12
3.3	Eentalig publicatietraject . . . . .	15
3.3.1	Problemen . . . . .	15
3.4	Meertalig publicatietraject . . . . .	18
3.4.1	Problemen . . . . .	21
3.5	Database-publicatietraject in een- of meertalig publicatietraject . . . . .	22
<b>4</b>	<b>Het vertaaltraject</b>	<b>23</b>
4.1	Vertalen . . . . .	23
4.2	De vertaler . . . . .	23
4.3	Het vertaalbureau . . . . .	24
4.4	Vertaaltrajecten . . . . .	24
4.4.1	Vertaaltraject van de vertaler . . . . .	24
4.4.2	Vertaaltraject van het vertaalbureau . . . . .	25
<b>5</b>	<b>Vertaalhulpmiddelen</b>	<b>29</b>
5.1	Tekstverwerker . . . . .	29
5.1.1	Invoeren en bewerken van tekst . . . . .	29
5.1.2	Inlezen en opslaan van tekst . . . . .	30
5.2	Termenbank . . . . .	30
5.3	Vertaalgeheugen . . . . .	31
5.4	Documentanalyse . . . . .	32
5.5	Alignment . . . . .	33
5.6	Vertaalomgeving . . . . .	33
5.7	Automatisch vertalen . . . . .	34
5.7.1	Verleden . . . . .	34
5.7.2	Verschillende systemen . . . . .	35
5.8	Automatische vertaling als zijeffect van database publicatie . . . . .	37

5.9	Overige hulpmiddelen . . . . .	37
5.9.1	Versiebeheer programmatuur . . . . .	38
5.9.2	Evaluatie van vertalingen . . . . .	38
<b>6</b>	<b>Algemene problemen bij het vertalen</b>	<b>39</b>
6.1	Problemen met karaktersets . . . . .	39
6.2	Problemen met opmaak van teksten . . . . .	40
6.3	Problemen met tabellen . . . . .	40
6.4	Problemen met ongebruikelijke bestandsformaten . . . . .	40
6.5	Problemen met vertaalgeheugens . . . . .	41
6.6	Problemen met terminologie . . . . .	42
6.7	Problemen met tags . . . . .	42
6.8	Problemen met communicatie . . . . .	43
<b>7</b>	<b>Interface tussen vertaal- en publicatietraject</b>	<b>45</b>
7.1	Communicatie . . . . .	45
7.2	Locatie . . . . .	46
7.2.1	Analyse van het document . . . . .	46
7.2.2	Vertalen van het document met een vertaalgeheugen . . . . .	47
<b>8</b>	<b>Aanbeveling voor verbeteringen</b>	<b>49</b>
8.1	Relatie met een vertaalbureau . . . . .	49
8.1.1	Vertaalhulpmiddelen . . . . .	49
8.2	Interne organisatie . . . . .	51
8.3	Afspraken met opdrachtgevers . . . . .	52
<b>9</b>	<b>Case</b>	<b>53</b>
9.1	Probleemstelling . . . . .	53
9.1.1	Huidige situatie . . . . .	53
9.1.2	Verwachtingen . . . . .	55
9.2	Organisatie en workflow . . . . .	56
9.3	Techniek . . . . .	57
9.3.1	Publicatietraject . . . . .	57
9.3.2	Programmatuur . . . . .	59
9.4	Kosten . . . . .	63
9.4.1	Opstartkosten . . . . .	63
9.4.2	Productiekosten . . . . .	64
<b>10</b>	<b>Toekomstvisie</b>	<b>65</b>
10.1	Vertaalhulpmiddelen . . . . .	65
10.1.1	Hardware . . . . .	65
10.1.2	Software . . . . .	67
10.2	Publicatie- en vertaaltraject . . . . .	73
10.2.1	Hulpmiddelen . . . . .	73
10.2.2	Vraag en aanbod . . . . .	75
<b>11</b>	<b>Conclusie</b>	<b>77</b>
<b>A</b>	<b>Evaluatie van vertaalhulpmiddelen</b>	<b>79</b>
	<b>Bibliografie</b>	<b>81</b>

# Voorwoord

Deze scriptie is een onderdeel van het resultaat van een stage bij Alfa▼*Base* voor de studie  $\alpha$ -informatica aan de Universiteit van Amsterdam. De stage duurde drie maanden. Naast deze scriptie is ook een verslag van een toekomstvisie en een aantal internet pagina's afgeleverd.

De scriptie beschrijft de resultaten van een onderzoek naar de combinatie van een vertaaltraject en een publicatietraject en de mogelijkheden tot automatisering daarvan. Bij een dergelijk complex traject komen taal, taaltechnologie, informatica en organisatie bij elkaar. Voor de studie  $\alpha$ -informatica is vooral de taal, taaltechnologie en informaticakant van het onderzoek belangrijk. Op die kant zal dus ook de nadruk gelegd worden, hoewel op sommige plaatsen ook gekeken wordt naar de organisatie van het totale traject.

Mijn begeleider bij de Universiteit van Amsterdam was Remko Scha. Bij het onderzoek heb ik veel informatie en ideeën gekregen van verschillende medewerkers van Alfa▼*Base*. Alan McGrane was daar mijn stagebegeleider. Hij heeft veel informatie over het publicatietraject gegeven en heeft mij gestuurd in dit onderzoek. Daarnaast hebben David Emmett, Mathilde Reedijk, Hans Scholte en Els van Trigt mij ook veel, vooral technische, informatie over het publicatietraject gegeven. Dennis Hamerling en Danielle Rossel van Berlitz International Translation Services hebben mij van informatie voorzien over het vertaaltraject en de gang van zaken in een vertaalbureau. Gerald Dennett heeft mij geholpen door een kopie van zijn major project report over vertaalgeheugens te sturen.



# Hoofdstuk 1

## Inleiding

Bij het publiceren van een document wordt een heel traject doorlopen, van het invoeren in een computer tot het drukken op papier of een ander medium: het publicatietraject. Dit traject wordt reeds jaren gebruikt. Door deze jarenlange ervaring zijn de meeste problemen binnen het traject al opgelost. Dit wil niet zeggen dat het traject vast staat en niet meer verandert. Nieuwe technieken maken het mogelijk sneller en beter te publiceren en nieuwe media maken nieuwe vormen van publiceren mogelijk.

Bij het vertalen van een document wordt eveneens een heel traject doorlopen, het vertaaltraject. Ook hier geldt dat dit al jaren gebeurt, dus ook hier zijn de meeste problemen al opgelost. Net als bij het publicatietraject maken ook hier nieuwe technieken het mogelijk sneller en beter te kunnen werken.

Problemen ontstaan als deze twee trajecten op elkaar aangesloten moeten worden. Dat wil zeggen, wanneer documenten in meerdere talen moeten worden gepubliceerd. Voor één enkel document is het traject nog te overzien, maar zeker wanneer de projecten groter worden en uit meerdere documenten en versies bestaan, ontstaan een aantal problemen.

Een aantal van deze problemen heeft te maken met de organisatie van beide trajecten. Omdat de trajecten elkaar gedeeltelijk overlappen is er op verschillende plaatsen de keus waar welke functies worden uitgevoerd. Elke mogelijkheid heeft zijn eigen eigenschappen en voor- en nadelen.

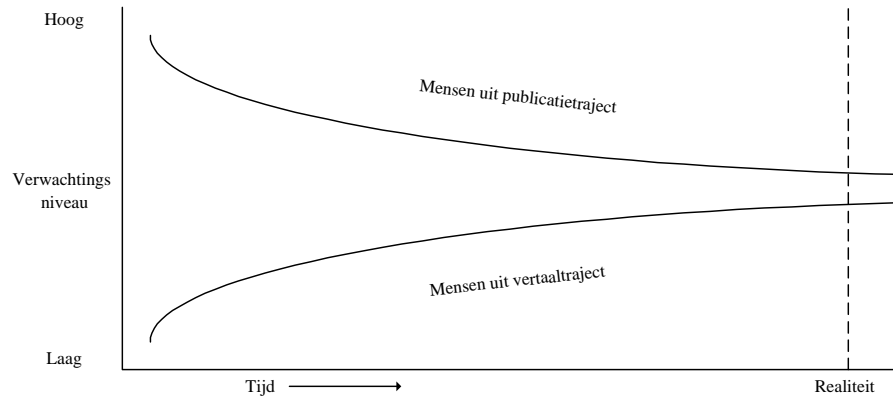
Een ander soort problemen bij de samenvoeging van beide trajecten heeft te maken met het overzicht van de informatie in beide trajecten. Bij het publicatietraject zonder het vertaaltraject blijft alle informatie binnen een bedrijf, namelijk het bedrijf dat de publicatie verzorgt. Als het vertaaltraject wordt toegevoegd, verplaatst informatie van publicatiebedrijf naar vertaalbedrijf en terug. Het overzicht van de locatie van bepaalde informatie kan daardoor verloren raken.

Een volgend soort probleem heeft te maken met de nieuwe technieken die in beide trajecten hun intrede doen. Als de trajecten los van elkaar staan, zijn bij de invoering van een nieuwe techniek slechts aanpassingen nodig in één traject. Door de combinatie van twee trajecten kan een nieuwe techniek nu beide trajecten tegelijkertijd aantasten. Bij het invoeren van een nieuwe techniek zijn in dit geval aanpassingen in het totale traject nodig.

Een laatste soort problemen heeft ook te maken met nieuwe technieken. Mensen uit het ene traject hebben vaak hoge verwachtingen van nieuwe tech-

nieken in het andere traject. [Rya93] geeft dit aan volgens de grafiek gegeven in figuur 1.1. Na een bepaalde tijd zullen de verwachtingen van beide partijen naar elkaar gegroeid zijn. Pas dan hebben beide partijen realistische verwachtingen van de nieuwe techniek.

**Figuur 1.1** verwachting en realiteit



Dit soort problemen wordt in deze scriptie nader beschreven en waar mogelijk worden oplossingen gegeven. Dit wil niet zeggen dat hier de *enige* oplossing gegeven wordt, in veel gevallen kunnen oplossingen alleen gevonden worden na uitgebreid overleg tussen de verschillende medewerkers uit beide trajecten. Ook kunnen de beste oplossingen per project net iets anders zijn.

Bij dit onderzoek wordt niet alleen gekeken naar de huidige problemen, maar er wordt ook rekening gehouden met de nodige aanpassingen voor de toekomst. Daarbij wordt aandacht besteedt aan de nieuwe technieken die in de toekomst beschikbaar zullen komen.

Deze scriptie is als volgt ingedeeld. In hoofdstuk twee wordt een beschrijving gegeven van de huidige situatie, gezien vanuit Alfa▼Base. Ook wordt daar aangegeven waarom veranderingen in de trajecten gewenst zijn. In hoofdstuk drie wordt het publicatietraject beschreven. Daar wordt onderscheid gemaakt tussen de verschillende vormen van het publicatietraject. Hoofdstuk vier bevat een beschrijving van het vertaaltraject. Hierbij worden verschillende aspecten binnen het traject beschreven, uitgaande van het vertaalbureau of de freelance vertaler. Daarna, in hoofdstuk vijf, volgt informatie over vertaalhulpmiddelen. De meest bekende en meest gebruikte soorten van vertaalhulpmiddelen worden daar behandeld. In hoofdstuk zes worden een aantal problemen van het publicatietraject beschreven, die voortkomen uit toevoeging van een vertaaltraject aan het publicatietraject. Bij het vertaaltraject ontstaan ook problemen. Ook deze problemen worden beschreven in hoofdstuk zes. Hoofdstuk zeven beschrijft de interface tussen het vertaal- en publicatietraject en in hoofdstuk acht worden aanbevelingen voor verbeteringen aangedragen. Deze aanbevelingen hebben betrekking op het publicatietraject en het vertaaltraject. In hoofdstuk negen wordt als voorbeeld een case behandeld. Deze case is een wat versimpelde versie van een praktijkgeval, omdat de beschrijving van het volledige praktijkgeval onnodige complexiteit met zich meebrengt. Hoofdstuk tien bevat een toekomstvi-

sie. Deze toekomstvisie is een samenvatting van de tekst die is beschreven in het aparte toekomstvisie-rapport. Op basis van deze vindingen kan het publicatie- en vertaaltraject aangepast worden, zodat zo goed mogelijk geanticipeerd kan worden op veranderingen in de toekomst. Als laatste wordt in hoofdstuk elf een conclusie van het totale onderzoek gegeven.





## Hoofdstuk 2

# Huidige situatie

In dit hoofdstuk wordt de huidige situatie bij Alfa▼*Base* beschreven. Er wordt een algemeen beeld geschetst waardoor een algemene context wordt beschreven. Er wordt hier niet diep op bepaalde processen ingegaan, dat zal in volgende hoofdstukken gebeuren.

Als eerste zal een overzicht gegeven worden van de verschillende categorieën van klanten van Alfa▼*Base*. Hiermee wordt duidelijk uit welke hoek de verschillende opdrachten komen en wat het soort documenten is dat gepubliceerd en vertaald moet worden.

Daarna wordt gekeken naar de problemen binnen de combinatie van het publicatie- en vertaaltraject zoals dat nu plaats vindt. De huidige trend binnen Alfa▼*Base* is om de trajecten verder te automatiseren en mogelijkheden te integreren tot één product. Er wordt gekeken naar de problemen die Alfa▼*Base* verwacht met deze trend op te lossen.

### 2.1 Situaties in de praktijk

Alfa▼*Base* deelt de klanten op in drie verschillende groepen. Deze indeling is gebaseerd op de verschillende soorten publicatie.

**Publishers** Het gaat hier om *zelfstandige externe* publicaties. Dit zijn publicaties die overwegend voor extern gebruik zijn en een op zichzelf staand product zijn. De klanten in deze groep hebben weinig of niets te vertalen. Het publiceren is hier het meest belangrijk. Alfa▼*Base* kan het publicatietraject uitvoeren, maar er zal meestal geen vertaaltraject plaatsvinden.

**Corporate publishers** In deze groep gaat het om *ondersteunende externe* publicaties. Dit zijn publicaties voor extern gebruik, die een ondersteunende functie hebben bij een apart product. Het gaat hier bijvoorbeeld om publicaties van handleidingen van een fax of een onderhoudshandleiding van een printer. Dit soort informatie zal vaak in veel verschillende talen gepubliceerd moeten worden. Deze bedrijven willen gebruik kunnen maken van een traject dat uit een publicatietraject gecombineerd met een vertaaltraject bestaat. Voorbeelden van dit soort bedrijven zijn Canon en PTT.

**Other** De publicaties die voor deze groep gepubliceerd worden zijn *interne* publicaties. Voorbeelden van interne publicaties zijn nieuwsbrieven en PR informatie. Deze informatie hoeft meestal niet of in slechts een klein aantal talen vertaald te worden. Vertalingen zullen nodig zijn voor internationale bedrijven die interne informatie in verschillende talen beschikbaar willen stellen. TNO en Shell zijn voorbeelden van dit soort bedrijven.

Voor de combinatie van het vertaaltraject en het publicatietraject zijn bedrijven in categorie 2 het meest interessant, hoewel bedrijven van categorie 3 soms ook projecten met vertalingen kunnen hebben.

## 2.2 De koppeling tussen het vertaal- en publicatietraject

Alfa▼Base heeft nu enkele projecten waarbij publicatie en vertaling met elkaar verweven zijn. Deze projecten hebben allemaal speciale eigenschappen die het traject ingewikkelder maken dan het algemene geval. Hierdoor zijn de projecten niet goed te gebruiken als voorbeelden voor het algemene geval. Om een goed overzicht te kunnen geven zullen we hier verder uitgaan van het algemene geval.

Als we kijken naar het publicatietraject gecombineerd met een vertaaltraject zien we dat op verschillende gebieden veranderingen ten opzichte van het eenzijdig publicatietraject nodig zijn. De grootste veranderingen zijn nodig in het publicatietraject, maar ook in het vertaaltraject zijn veranderingen noodzakelijk. Sommige veranderingen hebben te maken met het oplossen van problemen en andere veranderingen kunnen voor verbeteringen van het traject zorgen.

Het grootste gedeelte van het publicatietraject is geautomatiseerd. Dit heeft veel voordelen: de opmaak bijvoorbeeld kan automatisch gebeuren en documenten zijn gemakkelijk op te slaan. Ook de meeste vertaalbureaus zijn bezig het vertaaltraject te automatiseren. Dit wil zeggen dat vertalers door de computer worden bijgestaan tijdens het vertalen, maar ook kostenschattingen en kostenberekeningen worden meer en meer door de computer uitgevoerd.

De overgang tussen het vertaaltraject en het publicatietraject levert tot op heden nog problemen op. De beide trajecten zijn nog niet goed op elkaar afgestemd, waardoor de vertaler vaak niet goed met de aangeleverde documenten kan werken en de vertaalde documenten naderhand niet goed meer in het publicatietraject verwerkt kunnen worden. De verwachte oplossing voor dit probleem is het automatiseren en beter afstemmen van deze interface.

### 2.2.1 Oplossen van problemen

Bij een groot project waarbij een of meerdere documenten in veel talen gepubliceerd moet worden, ontstaan enkele organisatorische problemen. Een aantal van dat soort problemen komt in elk groter project voor, dus die zullen we hier niet beschrijven, maar een aantal heeft direct te maken met de aard van het project: het publiceren van documenten en hun vertalingen.

Een groot project heeft meerdere medewerkers, die ook nog eens op verschillende locaties kunnen werken. Hierdoor kan de communicatie tussen de verschillende medewerkers een probleem zijn.

Er zijn in principe twee soorten communicatie:

- *Communicatie waarbij organisatorische informatie wordt doorgegeven.* Dit is informatie die gaat over het traject zelf. Hierbij valt te denken aan problemen die een vertaler heeft met betrekking tot het gebruik van terminologie. Hij zal contact op moeten nemen met de opdrachtgever voor een oplossing. Een ander voorbeeld is dat het vertaalbureau doorgeeft dat een tekst in bepaalde talen vertaald is, maar dat een aantal vertalingen nog niet klaar zijn. Door dit soort informatie is het mogelijk bij te houden of de verschillende stappen in het traject goed op elkaar aansluiten.
- *Communicatie waarbij de te bewerken documenten worden doorgegeven.* Het totale traject bestaat uit verschillende stappen die elk als invoer een document hebben (en eventueel andere informatie) en als uitvoer hetzelfde, maar bewerkte, document hebben.

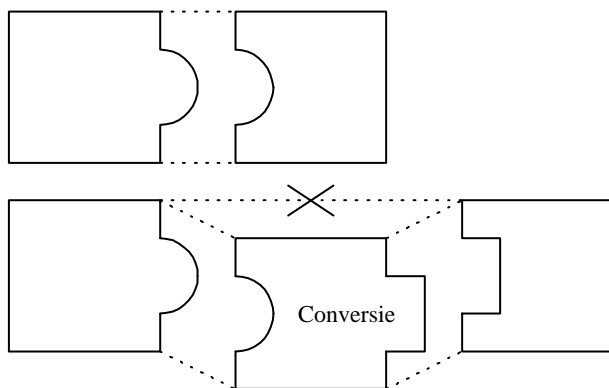
De eerste soort communicatie is een onderdeel van een algemeen project en komt dus ook in alle grotere projecten voor. Onze interesse gaat hier niet naar uit, omdat deze problemen niet direct te maken hebben met de combinatie van vertaal- en publicatietraject.

De tweede soort van communicatie kan meerdere problemen opleveren. Ten eerste moet elke stap in het totale traject de uitvoer van de vorige stap als invoer kunnen verwerken. Problemen hierbij hebben over het algemeen te maken met bestandsformaten. Het uitvoer bestandsformaat van de ene stap in het traject moet hetzelfde zijn als het invoer bestandsformaat van de volgende stap in het traject. Als dat niet het geval is, moet een conversiestap worden tussengevoegd. Om dit duidelijk te maken kunnen de afzonderlijke stappen als puzzelstukjes gezien worden. De ene stap (het ene puzzelstukje) moet op de volgende stap (het volgende puzzelstukje) passen. Als dat niet het geval is, moet een andere stap (een ander stukje) tussengevoegd worden. Zie figuur 2.1.

---

**Figuur 2.1** conversiestap

---



---

Ten tweede is het belangrijk dat bijgehouden wordt waar welke informatie aanwezig is. De informatie bestaat onder andere uit de te publiceren documenten, maar ook uit informatie als vertaalgeheugens, termenbanken en andere informatie die van belang is bij bepaalde stappen in het traject. Wanneer er geen duidelijk overzicht van de locatie van informatie is, zal het hele traject snel vertraging oplopen.

Dit zijn kort samengevat de grote problemen die bij de combinatie van de beide trajecten voorkomen. Alfa▼Base verwacht deze problemen op te kunnen lossen door de trajecten verder te automatiseren.

### **Automatiseren**

Het automatiseren van de communicatie kan helpen deze problemen op te lossen. Automatisering kan een goed communicatiekanaal opleveren, zodat communicatie gemakkelijker en sneller wordt. Wanneer gekozen wordt voor een open, elektronisch bestandsformaat, is het gemakkelijk de documenten door te geven naar de volgende stap in het traject.

Het grootste voordeel van goede communicatie is dat bijgehouden kan worden waar welke informatie aanwezig is. De verschillende documenten (en eventueel de verschillende versies van de documenten) kunnen gevolgd worden in elke stap in het traject. Hierdoor ontstaat een beter overzicht van de status van het project en kan er beter worden ingesprongen op plotselinge problemen die kunnen ontstaan.

De vraag is in hoeverre deze verwachtingen in de praktijk worden gehaald.

### **2.2.2 Verbeteringen door automatisering**

Het automatiseren heeft naast het oplossen van de bestaande problemen nog een ander, belangrijk voordeel, namelijk een financiële winst. Deze winst komt voort uit twee verschillende hoeken.

Voordat een traject geautomatiseerd kan worden moet het eerst gestandaardiseerd worden. Bij een gestandaardiseerd traject kan de tijdsduur en kosten van het traject beter geschat worden. Wanneer dit gestandaardiseerde traject geautomatiseerd wordt kan het traject soepeler verlopen. Er kan sneller op problemen worden ingesprongen en er kunnen betere plannings gemaakt worden, omdat het totale traject beter gevolgd kan worden. Dit zal, indien goed gebruikt, een kostenbesparing op kunnen leveren.

Omdat na het standaardiseren en automatiseren het traject soepeler verloopt zal het zeer waarschijnlijk ook sneller kunnen verlopen. Nu worden nog vaak problematische gevallen met de hand opgelost, maar als het traject geautomatiseerd is, zal meer en meer zonder ingreep van mensen kunnen gebeuren en is minder handwerk nodig. Dit levert een tijdswinst op. Deze tijdswinst zorgt niet alleen voor een lagere kostprijs, maar ook een betere service, omdat de meeste opdrachtgevers hun publicaties zo snel mogelijk in verschillende talen geleverd willen hebben. Het publiceren van documenten in verschillende talen is een complex traject dat veel tijd kost. Als het traject minder tijd in beslag neemt is dat een groot voordeel voor de opdrachtgevers.

### **Problemen**

Het publiceren in meerdere talen is een dermate complex traject, dat niet het hele traject in een keer geautomatiseerd kan worden. Door automatisering van bepaalde stappen in het traject, kunnen subtiele problemen ontstaan in volgende stappen.

Een ander probleem is dat bij het automatiseren verschillende hulpmiddelen nodig zijn, waarvan sommige nog niet beschikbaar of van voldoende niveau

zijn. Voorbeelden hiervan zijn conversieprogrammatuur en versiebeheer programmatuur.

Een volgend probleem is dat het totale traject niet alleen bij één, maar bij meerdere bedrijven plaats vindt: een bedrijf dat de publicatie verzorgt en een bedrijf dat de vertalingen verzorgt. Wanneer het gehele traject geautomatiseerd moet worden is het dus van belang dat de relatie tussen de bedrijven zeer goed is. De afzonderlijke trajecten moeten zo goed mogelijk op elkaar afgestemd worden. Dit is alleen mogelijk wanneer goede afspraken gemaakt worden en een goede communicatie aanwezig is.



## Hoofdstuk 3

# Het publicatietraject

In dit hoofdstuk zal het publicatietraject worden beschreven. Als eerste wordt beschreven wat publiceren is, daarna wordt ingegaan op het database-publicatietraject. Daarna wordt het eentalig publicatietraject beschreven. Als laatste wordt dit traject dan uitgebreid tot een meertalig publicatietraject.

### 3.1 Publiceren

Het overbrengen van informatie aan andere mensen kan op verschillende manieren. Het is bijvoorbeeld mogelijk de informatie over te brengen met behulp van spraak, per e-mail, via de radio of in boekvorm.

Bij een publicatie<sup>1</sup> is er sprake van "het uitgeven van een werk". In dat geval vallen mogelijkheden als spraak, e-mail en radio af. De nadruk zal hier liggen op uitgaven in papiervorm en gedeeltelijk op elektronische uitgaven (in de vorm van bijvoorbeeld cd-rom, diskette en internet pagina's).

Het publiceren van informatie kan gezien worden als een proces waarbij informatie de invoer is en de publicatie de uitvoer. Dat proces bestaat uit verschillende stappen die elk de informatie naar een andere vorm omzet zodat deze de invoer kan zijn voor de volgende stap in het proces.

Hoe het traject er precies uit ziet is afhankelijk van verschillende eisen aan de uiteindelijke publicatie. Hierbij is bijvoorbeeld te denken aan de vorm van de informatie (als invoer van het traject) en de vorm van de publicatie (de uitvoer van het traject). Het publicatietraject ziet er ook anders uit als er in meerdere talen uitgegeven moet worden dan bij de uitgave in een enkele taal.

In de volgende sectie wordt een overzicht geven van het database-publicatietraject. Daarna wordt een zo algemeen mogelijk publicatietraject beschreven waarbij gepubliceerd wordt in één taal. Daarna zal het meertalig publicatietraject behandeld worden.

---

<sup>1</sup>We zullen hier een verschil maken tussen een publicatie, zijnde het uitgegeven werk; een document, zijnde de totale, te publiceren informatie; en een tekst, zijnde de tekst in een document.

## 3.2 Database-publicatietraject

Als eerste in een publicatietraject moet een document gemaakt worden. Dit is een stap waar Alfa▼*Base* meestal niets mee te maken heeft, er wordt immers een document ter publicatie aangeleverd. Wanneer het maken van een document goed geïntegreerd wordt in het publicatietraject kan dit bepaalde voordelen hebben. Daarom wordt hier eerst dit traject beschreven.

Het maken van een document kan op verschillende manieren.

**Schrijven** Bij het schrijven van een document is er iemand bezig een document te maken, waarbij gezocht wordt naar informatie in naslagwerken of andere informatiebronnen.

**Samenstellen (database publicatie)** Wanneer een document wordt samengesteld, is er iemand bezig stukken reeds bestaande tekst te selecteren en aan elkaar te “plakken”. In een database staan grote stukken document opgeslagen. Deze stukken document worden hergebruikt bij het creëren van een nieuw document.

**Combinaties van beide manieren** Bij combinaties van beide manieren valt te denken aan het maken van een document met stukken document die worden geschreven en stukken die samengesteld worden. In dit geval worden stukken bestaand document hergebruikt.

Dit is in principe een heel spectrum aan mogelijkheden. Als een document geschreven wordt wil dat meestal niet zeggen dat de volledige tekst bedacht wordt. Meestal worden delen uit andere documenten gebruikt. De “database” is dan in het hoofd van de schrijver aanwezig. Hij weet welk document het stuk tekst bevat dat hij wil hergebruiken.

Wanneer tekstdelen op een dergelijke manier hergebruikt worden, moeten soms delen tekst opnieuw ingetypt worden of opnieuw opgemaakt worden. De tekstdelen kunnen in verschillende stadia van ontwikkeling zijn. Het is dan dus niet zo gemakkelijk als bij “pure database-publicatie” om tekstdelen te hergebruiken. Om de teksten in verschillende stadia van ontwikkeling te kunnen gebruiken voor database-publicatie is verder onderzoek nodig.

Het schrijven van een document is altijd mogelijk, maar het samenstellen van een document kan veel voordelen hebben. De grote voordelen van het samenstellen van een document zijn

- De documentatie is gestandaardiseerd. Hierdoor kunnen veel stappen in het publicatie- en vertaaltraject automatisch gebeuren.
- De stukken documentatie kunnen meerdere malen hergebruikt worden. Een document kan sneller worden samengesteld omdat er geen tekst geschreven en gecontroleerd moet worden, er hoeft slechts eens stuk document geselecteerd te worden.

Het samenstellen van een document is helaas niet altijd mogelijk. Dit is afhankelijk van verschillende zaken.



**Wat voor soort document is het** Documenten die duidelijk gestructureerd zijn, bijvoorbeeld handleidingen of labels, zijn beter geschikt voor het samenstellen van een document dan bijvoorbeeld brochures of reclamemateriaal. Vooral wanneer het document technische informatie bevat kan het samenstellen van een document veel voordelen bieden.

**Hoeveel van het document kan hergebruikt worden** Wanneer een document geschreven is, kunnen soms grote delen hergebruikt worden bij het samenstellen van andere documenten.

**Moet het document vertaald worden** Als een document is samengesteld en al een keer is vertaald (bij de eerste keer dat de delen gebruikt werden) kan bijna automatisch een vertaling van het document gemaakt worden. Alle delen zijn immers al een keer vertaald.

**Is een database met informatie beschikbaar** Bij het samenstellen van een document moeten uiteraard wel de delen van de documenten beschikbaar zijn. Over het algemeen zullen delen van documenten in een database opgeslagen worden, zodat ze gemakkelijk terug te vinden zijn en tegelijkertijd kan extra informatie erbij opgeslagen worden, zoals vertalingen.

De manieren om een document samen te stellen zoals hier beschreven hebben voor- en nadelen. Een combinatie van de twee manieren kan soms enkele nadelen oplossen, maar kan zelf ook nadelen hebben. Ook hier is dit afhankelijk van de zaken zoals hierboven beschreven.

Als we combinaties van het schrijven en samenstellen van documenten even buiten beschouwing laten, kunnen we beide manieren tegenover elkaar zetten.

Het schrijven van een document	Het samenstellen van een document
Documenten worden gemaakt door alle tekst helemaal te schrijven.	Documenten worden gemaakt door het kiezen van stukken tekst.
Informatie kan in elke vorm aanwezig zijn.	Informatie moet in de database aanwezig zijn.
Nieuwe tekst kan direct verwerkt worden.	Nieuwe tekst moet speciaal behandeld worden, zodat het in de database komt te staan.
Documenten moeten elke keer vertaald worden.	Documenten kunnen automatisch vertaald worden als vertalingen in de database worden opgeslagen.
Voor steeds herhalende teksten met bijvoorbeeld informatie uit een database moeten elke keer met de hand gegevens veranderd worden.	Informatie uit de (gegevens) database kan direct in de tekst verwerkt worden, zodat een soort standaard bestand ontstaat dat ingevuld kan worden met gegevens uit de database.

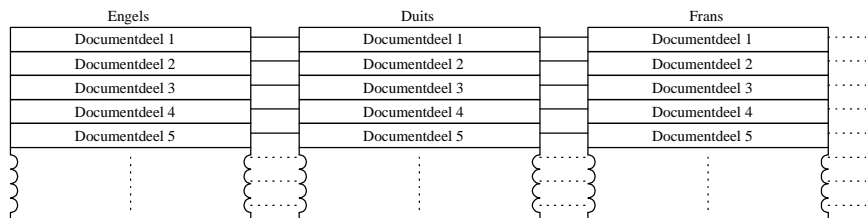
Beide mogelijkheden hebben voor- en nadelen. Wat precies voor- en nadelen zijn is afhankelijk hoe het systeem gebruikt wordt. Over het algemeen heeft het schrijven van een document het voordeel dat er zeer creatief gewerkt kan worden aan het document. De voordelen van het samenstellen van een document zijn dat een zeer gestandaardiseerd document verkregen wordt, dat bijvoorbeeld automatisch vertaald kan worden. Het grote nadeel van het schrijven van een document is dat elk document apart vertaald moet worden. Het nadeel van het

samenstellen van een document is dat een creatief schrijver zich in zijn werk belemmerd voelt.

Afhankelijk van bovengenoemde voor- en nadelen kan gekozen worden voor een database-publicatietraject. Hierbij hoeft verder niets te veranderen in de rest van het publicatietraject, maar het is wel mogelijk het publicatietraject zo in te richten dat vooral het vertalen van documenten efficiënter kan.

Omdat bij een (puur) database-publicatietraject alle teksten uit een database komen, is het mogelijk de vertalingen van de teksten bij de originelen in de database op te slaan. Wanneer de documenten goed in delen ingedeeld zijn, is het mogelijk ook de vertalingen te hergebruiken. Na het samenstellen van een tekst uit de database (in één taal), zijn direct vertalingen beschikbaar. Dit is mogelijk omdat de database een structuur heeft zoals beschreven in figuur 3.1. Elk veld dat gekozen is in een bepaalde taal, heeft een vertaling. Een nieuw document wordt dus een lijst van veldnamen. Afhankelijk van welke taal gekozen wordt, is het document beschikbaar in die taal.

**Figuur 3.1** structuur van database bij database-publicatie



Zoals het nu gebruikt wordt zijn er twee problemen bij het database-publicatietraject.

1. Bij het samenstellen van een document moeten verschillende tekstdelen geselecteerd worden waaruit het document zal bestaan. Er moet dus een goede gebruikersinterface zijn zodat tekstdelen gemakkelijk op te zoeken zijn. Uiteraard is de interface afhankelijk van de gebruikte informatie.

Als complete teksten worden gegenereerd vanuit een database is een goed zoekstelsel van belang. Bij het samenstellen van teksten uit een database moeten gemakkelijk bepaalde tekstdelen die in de database aanwezig zijn, gevonden kunnen worden. In de praktijk gaat dit meestal met behulp van een full-text zoekstelsel. Een beter stelsel zou zijn een full-text zoekstelsel dat zoekt op semantische informatie. Dat wil zeggen dat niet alleen op bepaalde woorden wordt gezocht, maar ook op woorden die wat betekenis betreft lijken op het te zoeken woord. Hierdoor kunnen ook teksten gevonden worden waar niet het specifieke zoekwoord instaat, maar een woord dat dezelfde of ongeveer dezelfde betekenis heeft.

Wanneer database publicatie alleen gebruikt wordt voor technische informatie in een document moet het bijvoorbeeld mogelijk zijn een selectie te maken van informatie in de database.

2. Er moet een duidelijke handelswijze afgesproken worden hoe er om wordt gegaan met tekstdelen die nog niet in de database aanwezig zijn. Over het algemeen bestaat een nieuw document niet volledig uit al bestaande tekstdelen, dus wanneer een nieuw tekstdeel in een document gebruikt wordt, moet dat op de een of andere manier toegevoegd worden aan de database. Wanneer dat niet gebeurt zijn ook bijvoorbeeld automatische vertalingen niet meer mogelijk. Een mogelijke oplossing voor dit probleem is te zorgen dat nieuwe tekstdelen door de schrijver toegevoegd worden aan de database. Op een bepaalde moment worden alle “nieuwe” tekstdelen uit de database gehaald. Daarna worden deze tekstdelen vertaald en ook aan de database toegevoegd. Hierna is de database weer volledig en zijn automatische vertalingen weer mogelijk.

### 3.3 Eentalig publicatietraject

Elke publicatie doorloopt ongeveer hetzelfde traject, hoewel bepaalde stappen iets anders kunnen zijn. Het algemene traject ziet er uit als weergegeven in figuur 3.2.

Het traject begint met de informatie die uitgegeven moet gaan worden. Deze informatie wordt in de computer ingevoerd, waarna een conversiestap plaats vindt. Na de conversie kan de uitvoer gegenereerd worden. Eventueel kan daarna nog een afwerkingstap gemaakt worden.

In figuur 3.3 staan de verschillende stappen ingevuld met mogelijkheden zoals die over het algemeen bij Alfa▼*Base* voorkomen. Dit wil niet zeggen dat bij elk project precies die stappen aanwezig zijn. Elk project heeft een iets andere invulling van de stappen en elk project heeft zijn eigen probleemgevallen.

Elke conversiestap heeft zijn eigen hulpmiddelen. Bij de conversie van informatie op papier naar ongestructureerd ASCII wordt gebruikt gemaakt van een scanner en OCR software<sup>2</sup>. Hierdoor ontstaat ongestructureerd ASCII dat handmatig van opmaakcodes (met behulp van bijvoorbeeld WordPerfect) voorzien wordt. Gestructureerd ASCII wordt met behulp van scripting talen als Perl en Omnimark automatisch voorzien van opmaakcodes zodat de (opgemaakte) informatie verwerkt kan worden door de zet- of opmaakprogrammatuur als XYVision of Quark XPress. XYVision en Quark XPress worden gebruikt om de uitvoer te genereren. Verder wordt voor de handmatige conversie van SGML bestanden de SGML editor Adept gebruikt. Voor het tekenen en bewerken van tekeningen en foto's wordt Photoshop en Illustrator gebruikt.

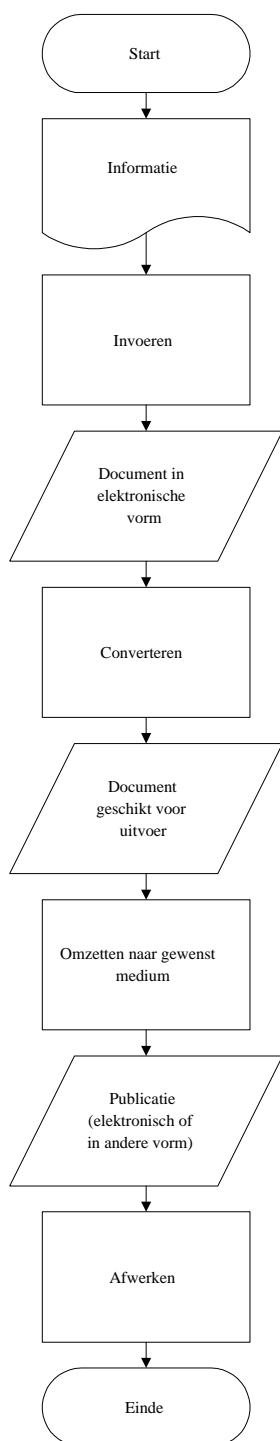
#### 3.3.1 Problemen

De voornaamste problemen binnen dit traject hebben te maken met het converteren. Bijna elke conversiestap kan problemen opleveren. Wanneer bijvoorbeeld informatie op papier wordt aangeleverd, moet de informatie eerst in de computer gebracht worden met behulp van een scanner en OCR software. Hierbij ontstaan fouten die verbeterd moeten worden.

Het feit dat niet alle bestandsformaten overdraagbaar zijn is een ander probleem. Bij conversie gaat opmaak verloren of gaan tags<sup>3</sup> verloren. Dit soort

<sup>2</sup>Optical Character Recognition, tekstherkenningssoftware.

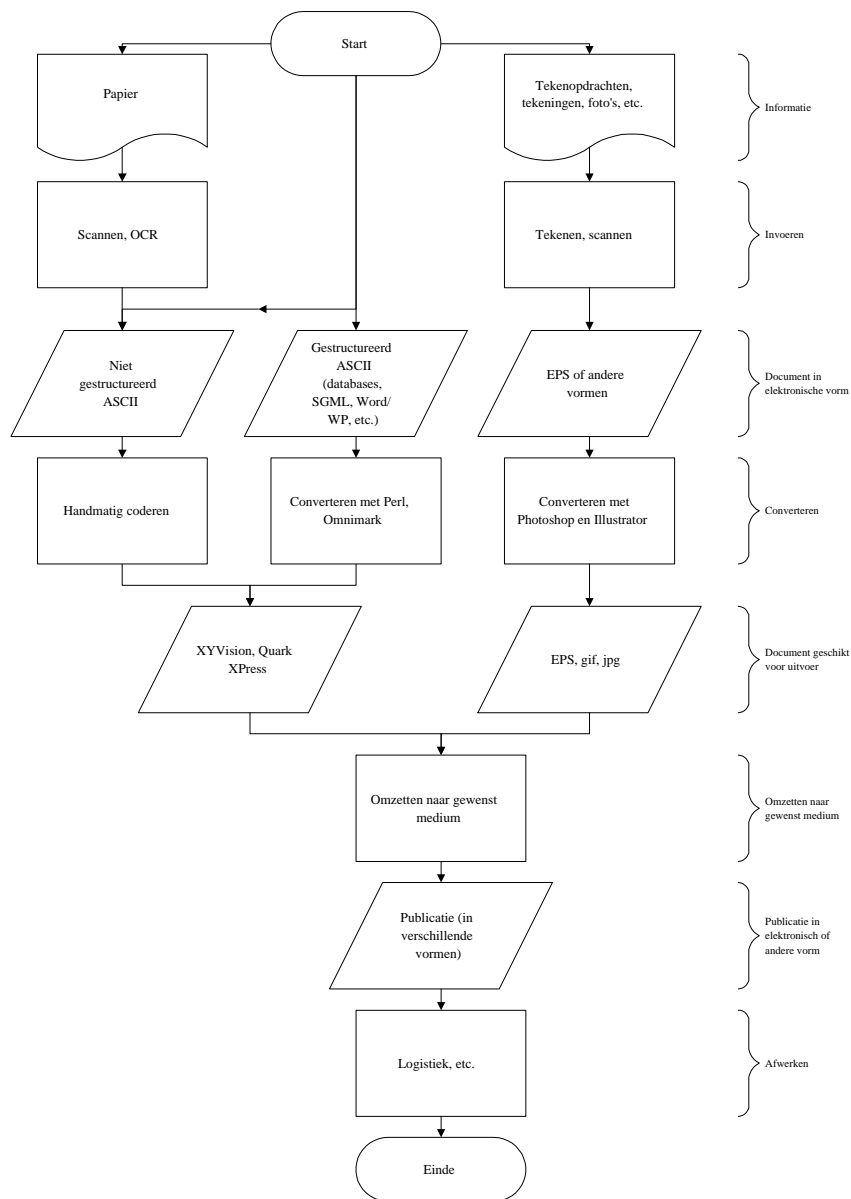
<sup>3</sup>Tags zijn, over het algemeen leesbare, codes in een document, waarmee structurele infor-

**Figuur 3.2** eentalig publicatietraject

---

**Figuur 3.3** publicatietraject bij Alfa▼Base
 

---



fouten moet naderhand weer met de hand aangepast worden. Vaak zijn echter de bestandsformaten niet leesbaar voor de mens, zodat specifieke software gebruikt moet worden.

Ook kan de oorspronkelijke invoer fouten bevatten waardoor automatische conversie niet altijd mogelijk is.

### 3.4 Meertalig publicatietraject

Het verschil tussen een eentalig publicatietraject en een meertalig publicatietraject is het aantal talen waarin de publicatie geproduceerd wordt. Bij het eentalig publicatietraject wordt de publicatie in één taal geproduceerd, namelijk de taal waarin de oorspronkelijke informatie wordt aangeleverd. Bij een meertalig publicatietraject wordt de informatie tijdens het publicatietraject vertaald naar andere talen. (Een publicatietraject waarbij informatie in meerdere talen wordt aangeleverd kan gezien worden als meerdere eentalige publicatietrajecten.)

Omdat het vertalen tijdens het publicatietraject plaats vindt, ziet het verloop van het traject er iets anders uit dan bij het eentalig publicatietraject. Het meertalig publicatietraject is weergegeven in figuur 3.4. Dit is hetzelfde als bij het eentalig publicatietraject alleen is een vertaalstap toegevoegd. Uiteraard is de informatie na de vertaalstap aanwezig in meerdere talen.

Het meertalig publicatietraject is iets complexer dan is weergegeven in figuur 3.4. De vertaalstap zelf is niet een enkele stap, maar bestaat uit meerdere stappen. Verder is er na het opmaken van de informatie een terugkoppeling naar de vertaler zodat de opgemaakte informatie zelf gecontroleerd kan worden.

De tekst wordt opgemaakt met behulp van opmaakcodes in de tekst. Het verzorgen van de opmaak kan bij verschillende stappen in het traject.

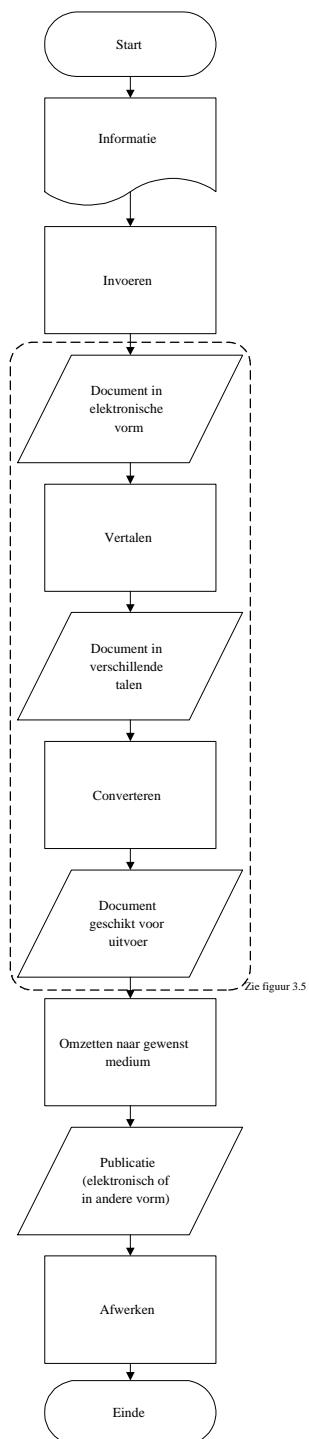
- *De opmaakcodes worden toegevoegd voor de vertaalstap.* Hierbij moet de vertaler de opmaakcodes in de vertaling overbrengen, zodat de vertaling automatisch hetzelfde is opgemaakt als het origineel.
- *De opmaakcodes worden toegevoegd na de vertaalstap.* De vertaler hoeft dan niet direct rekening te houden met de opmaak. Naderhand moet de vertaler controleren of het opmaken van de vertaling goed plaats heeft gevonden.

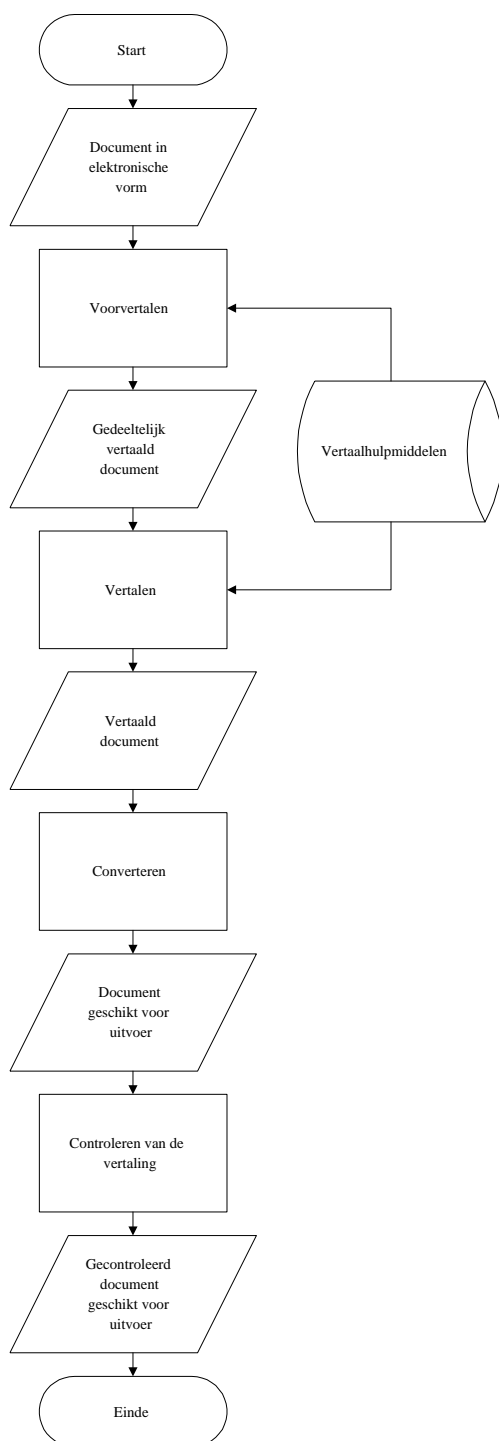
Vaak zullen beide mogelijkheden gekozen worden. De vertaler probeert dan zoveel mogelijk van de bestaande opmaak in de vertaling te brengen. Na het vertalen wordt de opmaak gecontroleerd en aangepast waar nodig.

Figuur 3.5 geeft een completer beeld van de vertaalstap. Dit figuur is in te passen binnen de stippellijn in figuur 3.4.

Het voorvertalen gebeurt bij Alfa▼Base met behulp van Transit, een vertaalprogramma dat gebruik maakt van een vertaalgeheugen. Na deze stap is de informatie gedeeltelijk vertaald. De delen van de informatie die nog niet vertaald zijn, worden opgestuurd naar een vertaalbureau dat de niet vertaalde delen vertaalt. Na deze stap is de informatie volledig vertaald. De informatie wordt daarna bij Alfa▼Base opgemaakt en weer opgestuurd naar het vertaalbureau waar de informatie gecontroleerd wordt. Dit wordt gedaan om bijvoorbeeld

matie aangegeven kan worden. Deze structurele informatie wordt meestal naderhand omgezet in opmaakinformatie.

**Figuur 3.4** publicatietraject inclusief vertaalstap

**Figuur 3.5** vertaaltraject binnen publicatietraject



verkeerd afgebroken woorden te verbeteren. Daarna is de informatie klaar voor uitvoer.

Het vertalen van tekst bij illustraties werkt net iets anders. Als eerste wordt de tekst uit de illustraties gehaald en worden de coördinaten onthouden. Daarna wordt het illustratie met de tekst en de coördinaten in een database opgeslagen. De tekst wordt vertaald net als in het gewone vertaaltraject en naderhand wordt de vertaalde tekst op de goede plaats in het illustratie neer gezet.

Het vertaaltraject met afbeeldingen zoals hier geschetst is uiteraard slechts voor één taal. Wanneer in meerdere talen uitgegeven moet worden, zal voor elk van de talen een dergelijk traject doorlopen moeten worden.

### 3.4.1 Problemen

Het meertalig publicatietraject is veel complexer dan het eentalig publicatietraject. Bij dit traject kunnen dezelfde problemen als bij het eentalig publicatietraject zich voordoen, maar naast die problemen heeft dit traject ook zijn eigen problemen.

#### Versiebeheer

Tijdens het traject zijn, zeker wanneer in veel talen vertaald moet worden, zeer veel verschillende versies van de informatie aanwezig. Verschillende versies zijn: het origineel, de gedeeltelijk vertaalde versies (in de verschillende talen), de vertaalde versies en de verbeterde versies. Ook komt het vaak voor dat terwijl een bepaalde versie reeds bewerkt is, een nieuwe versie geleverd wordt. Wanneer geen goed versiebeheer aanwezig is, is het snel onduidelijk welke versie zich waar bevindt.

#### Kennis van de taal

Mensen moeten werken met informatie die niet in hun moedertaal is geschreven. Het is hierdoor moeilijk, zo niet onmogelijk, bepaalde operaties op de informatie uit te voeren. Het is bijvoorbeeld moeilijk aan te geven waar woorden afgebroken moeten worden. Ook bij de opmaak van de informatie kunnen problemen ontstaan. Vaak moet bepaalde tekst op een bepaalde bladzijde staan, zodat de referenties goed blijven. Een vertaling kan echter meer plaats innemen dan bij het origineel. Hetzelfde probleem ontstaat bijvoorbeeld bij tabellen en illustraties.

#### Problemen met karaktersets

Er ontstaan problemen bij verschillende karakters die in talen voor kunnen komen. Aangezien sommige talen veel speciale karakters gebruiken moet daar rekening mee worden gehouden met het gebruikte bestandsformaat. Andere problemen die hierbij ontstaan zijn wanneer informatie in “vreemde” karaktersets veranderd moet worden. Omdat deze “vreemde” karakters niet op het toetsenbord aanwezig zijn, moeten hier andere hulpmiddelen gebruikt worden.

### **Kwaliteit van de vertaling**

Er wordt vaak vanuit gegaan dat de vertalingen “goed” zijn, hoewel dit niet het geval hoeft te zijn. Het is moeilijk te zeggen wat een goede vertaling is. Dat is onder andere afhankelijk van wat voor eisen aan de vertaling gesteld worden. Aan verschillende soorten informatie kunnen andere eisen gesteld worden. Brochures zijn bijvoorbeeld niet of erg moeilijk te vertalen vanwege cultuur afhankelijkheden, terwijl handleidingen zeer goed te vertalen zijn. Verder zijn de gestelde eisen afhankelijk van het doel van de informatie. Bij handleidingen is het minder erg als deze tik- of taalfouten bevat, zo lang het voor de gebruiker maar duidelijk is wat er moet gebeuren. Het is erger als er fouten zitten in een brochure, omdat een brochure een bepaald beeld geeft van een bedrijf of product.

## **3.5 Database-publicatietraject in een- of meertalig publicatietraject**

Het database-publicatietraject overlapt het een- of meertalig publicatietraject voor een gedeelte. Hoe het database-publicatietraject precies het publicatietraject overlapt is afhankelijk van de uitvoer van de database-publicatie. Als het document, na het selecteren van tekst uit de database, uit niet gestructureerde tekst in één taal bestaat, moet in elk geval de opmaak nog toegevoegd worden en eventueel het document nog vertaald worden. Het is bijvoorbeeld ook mogelijk dat na het selecteren van de tekstdelen een document in meerdere talen volledig opgemaakt beschikbaar is. Afhankelijk van de staat waarin het samengestelde document zich bevindt, past het database-publicatietraject op het publicatietraject.

## Hoofdstuk 4

# Het vertaaltraject

In dit hoofdstuk wordt een beschrijving van het vertaaltraject gegeven. Als eerste wordt een korte beschrijving gegeven van het vertaaltraject waarna wordt ingegaan op de rol van de vertaler binnen dit traject. Daarna wordt ingegaan op de functie van het vertaalbureau. Vervolgens worden twee verschillende vertaaltrajecten beschreven.

### 4.1 Vertalen

Het vertaaltraject kan worden beschreven als een proces waarbij een document in de brontaal als invoer en het document in de doeltaal als uitvoer geldt. Wanneer een document vertaald moet worden, zijn er twee mogelijkheden. Er wordt direct contact gezocht met een freelance vertaler of het document wordt opgestuurd naar een vertaalbureau. Als een vertaler of vertaalbureau gevonden is, kan het document in de brontaal naar de vertaler of het vertaalbureau en na het vertalen komt het vertaalde document terug.

Als er op deze manier naar het vertaaltraject gekeken wordt, zou er geen verschil zijn tussen een freelance vertaler en een vertaalbureau. Wat de vertaling zelf betreft is dat er ook niet, maar verschillen zijn wel te vinden bij ondersteunende diensten naast het vertalen.

### 4.2 De vertaler

Een vertaler kan op twee verschillende manieren worden ingezet, als freelance vertaler of als vertaler in dienst van het vertaalbureau. In principe werken ze allebei op dezelfde manier, het enige verschil is dat de freelance vertaler een opdracht direct van een opdrachtgever kan krijgen, terwijl een vertaler in dienst van een vertaalbureau de opdracht via het vertaalbureau krijgt.

Een vertaler is over het algemeen gespecialiseerd in het vertalen naar één vreemde taal. Als een document in meerdere talen vertaald moet worden, zijn dus meerdere vertalers nodig, die elk naar één andere taal vertalen.

Freelance vertalers zullen meestal twee verschillende soorten opdrachten krijgen:

1. kleine, eenmalige opdrachten; hierbij kan gedacht worden aan vertalingen van wetenschappelijke artikelen, menukaarten en persoonlijke correspondentie,
2. grotere eventueel zich herhalende opdrachten gecoördineerd vanuit een vertaalbureau.

Vertalers die bij een vertaalbureau werken krijgen uiteraard de opdrachten direct van het vertaalbureau. Dit kunnen zeer uiteenlopende opdrachten zijn.

### 4.3 Het vertaalbureau

Een vertaalbureau is, in tegenstelling tot wat de naam zegt, over het algemeen niet een bureau waar mensen vertalen. Een vertaalbureau heeft meer de taak het vertaaltraject te coördineren. Dit wil zeggen, vertalers voor een opdracht te zoeken en de verschillende versies van het document te beheren. Hoewel sommige vertaalbureaus, afhankelijk van de grootte en hoeveelheid van te vertalen teksten, vertalers direct in dienst hebben, maken de meeste vertaalbureaus gebruik van freelance vertalers.

### 4.4 Vertaaltrajecten

Beide mogelijkheden tot het vertalen van documenten hebben hun eigen traject. Bij de vertaler (freelance of in dienst van een vertaalbureau) vindt het eigenlijke vertaaltraject plaats, terwijl het vertaalbureau de organisatie van het vertaaltraject behandelt. Het vertaaltraject van de vertaler is dus een onderdeel van het vertaaltraject van het vertaalbureau.

#### 4.4.1 Vertaaltraject van de vertaler

In figuur 4.1 is het vertaaltraject van de vertaler weergegeven. Er zijn verschillende mogelijkheden.

- De vertaler vertaalt het document met de hand, zonder hulp van vertaalhulpmiddelen.
- De vertaler vertaalt het document met hulp van vertaalhulpmiddelen, ongeacht of de vertaalhulpmiddelen iets opleveren.
- De vertaler analyseert het document, maar concludeert dat vertaalhulpmiddelen niet of weinig helpen, dus vertaalt het document zonder hulp van vertaalhulpmiddelen.
- De vertaler analyseert het document en concludeert dat vertaalhulpmiddelen wel helpen, dus vertaalt met hulp van vertaalhulpmiddelen.
- De opdrachtgever geeft de opdracht dat vertaalhulpmiddelen gebruikt moeten worden.

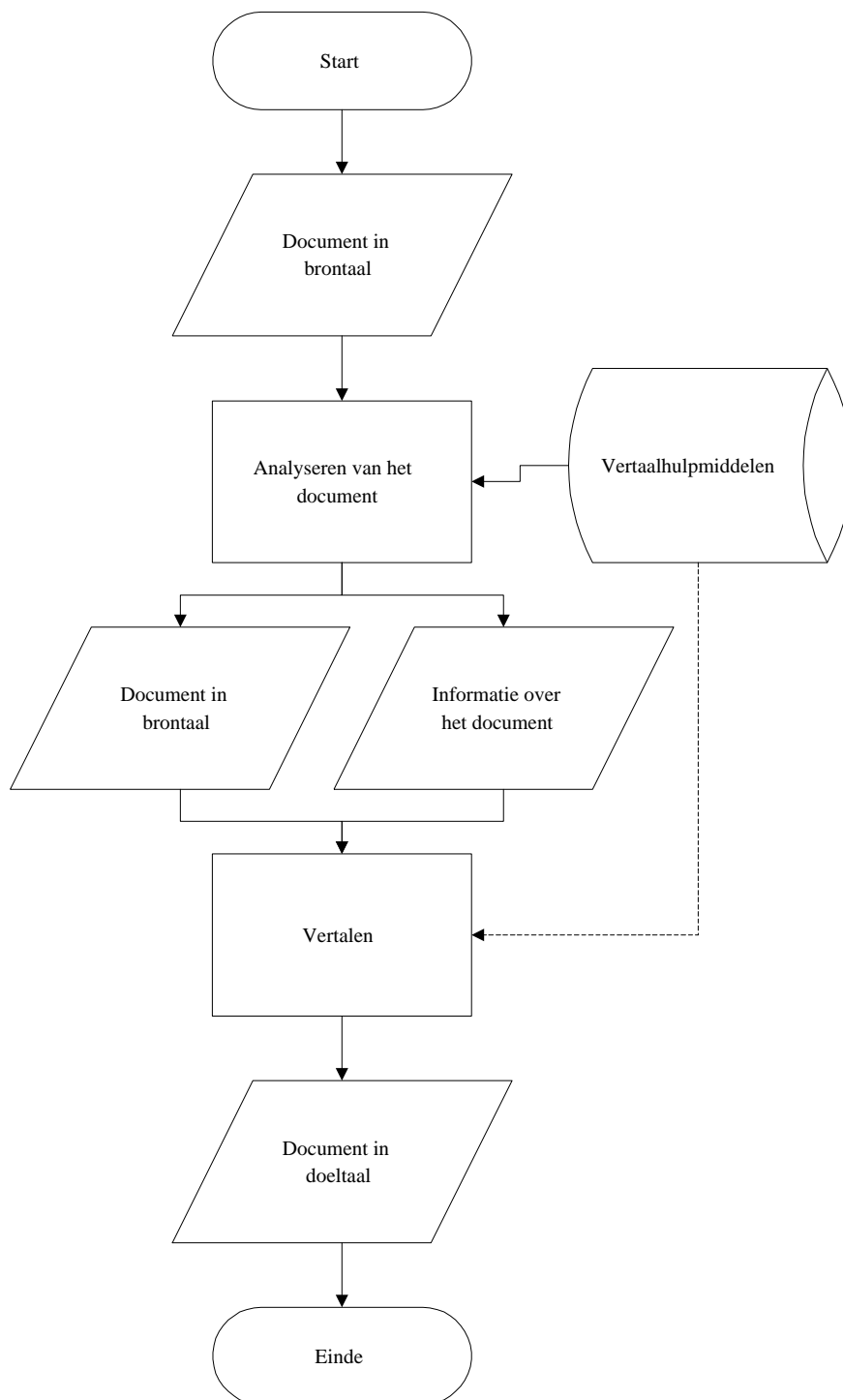
Afhankelijk van de analyse van het document kan de vertaler bepalen in hoeverre een vertaalhulpmiddel helpt bij het vertalen van het document. De uiteindelijke keuze hangt af van verschillende factoren.

- *De beschikbaarheid van een vertaalgeheugen-database.* Wanneer er geen vertaalgeheugen-database aanwezig is of met het document meegeleverd is en er niet gemakkelijk een vertaalgeheugen-database gegenereerd kan worden uit vorige vertalingen, zal zeer waarschijnlijk geen vertaalgeheugen gebruikt worden bij de vertaling.
- *Hoeveel 100% matches in het document voorkomen (gebaseerd op de vertaalgeheugen-database).* Als een vertaalgeheugen-database aanwezig is en er veel zinnen in het geheugen voorkomen in het document zal waarschijnlijk wel gebruik gemaakt worden van het vertaalgeheugen. Dit kan bijvoorbeeld het geval zijn wanneer verbeteringen aangebracht zijn in een reeds eerder vertaalde tekst.
- *De hoeveelheid zinnen die meerdere malen in het document voorkomt.* Als zinnen meerdere malen voorkomen in een document kan een vertaalgeheugen helpen. De zin hoeft dan slechts één keer vertaald te worden en bij de volgende keer dat de zin voorkomt wordt de vertaling aangeboden uit de vertaalgeheugen-database.
- *De beschikbaarheid van een termenbank.* Als de opdrachtgever een termenbank van goede kwaliteit ter beschikking stelt, is het zeer gemakkelijk deze te gebruiken. Wanneer geen termenbank beschikbaar is, kost het mogelijk te veel tijd deze eerst te produceren.
- *De hoeveelheid terminologie in het document.* Afhankelijk van de hoeveelheid terminologie in het document kan gekozen worden of er een termenbank gebruikt gaat worden en eventueel, wanneer geen termenbank beschikbaar is of er een gemaakt wordt.
- *Is er een vorige (vertaalde) versie van het document.* Uit vorige vertaalde versies van documenten is vrij gemakkelijk een vertaalgeheugen-database te genereren. Hierdoor kunnen, vooral bij verbeteringen binnen een document, grote stukken tekst automatisch vertaald worden. Alle niet veranderde tekst in het origineel zal namelijk automatisch vervangen worden.
- *Stelt de opdrachtgever de eis om bepaalde hulpmiddelen te gebruiken.* De opdrachtgever kan eisen stellen aan de gebruikte hulpmiddelen bij het vertaaltraject. Afhankelijk van de motivatie van het vertaalbureau of de opdrachtgever om bepaalde hulpmiddelen te gebruiken of niet te gebruiken kan worden besloten bepaalde hulpmiddelen bij het vertaaltraject in te zetten. Hoewel de opdrachtgever of Alfa▼Base meestal niet de kennis heeft over het vertalen zelf, kunnen zij beter het totale traject overzien. De vertaler kan alleen beslissen vanuit zijn eigen context.

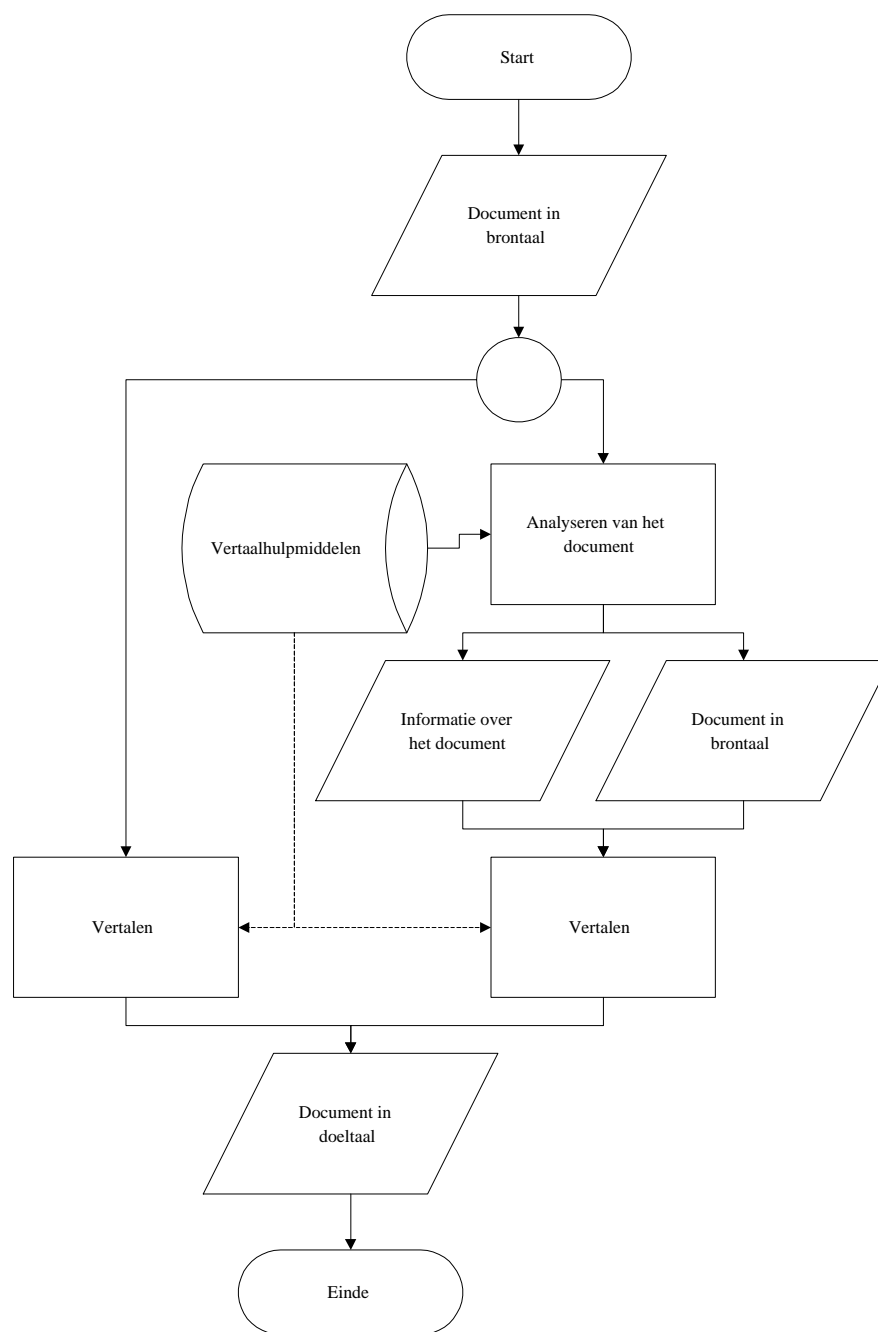
#### 4.4.2 Vertaaltraject van het vertaalbureau

In figuur 4.2 is het traject weergegeven dat een document bij een vertaalbureau doorloopt. Hier is te zien dat het vertaaltraject van de vertaler een onderdeel is van het traject van het vertaalbureau.

Als eerste vindt een analyse van het te vertalen document plaats. Op basis hiervan wordt gekozen welke vertaalhulpmiddelen gebruikt gaan worden en wat

**Figuur 4.1** vertaaltraject

de kosten van de vertaling worden. Daarna wordt het document naar de vertaler, freelance of bij het vertaalbureau zelf, gestuurd. Deze vertaalt het document, eventueel met behulp van vertaalhulpmiddelen. Als het document vertaald is komt het terug bij het vertaalbureau. Daar wordt het document teruggestuurd naar de opdrachtgever.

**Figuur 4.2** analyse van het te vertalen document



## Hoofdstuk 5

# Vertaalhulpmiddelen

Bij het vertaaltraject wordt gebruik gemaakt van verschillende hulpmiddelen. In dit hoofdstuk zullen de belangrijkste hulpmiddelen worden beschreven. Welke hulpmiddelen precies gebruikt worden bij een project is afhankelijk van verschillende factoren. Niet bij elk project worden alle hulpmiddelen gebruikt.

Als eerste zullen we de tekstverwerker beschrijven, daarna de termenbank, gevolgd door het vertaalgeheugen. Dit zijn hulpmiddelen die direct bij het vertalen gebruikt worden. Daarna wordt aandacht besteed aan documentanalyse en alignment hulpmiddelen. Deze hulpmiddelen zijn van een meer ondersteunende aard. De vertaalomgeving, die daarna wordt beschreven, is een integratie van de eerder beschreven hulpmiddelen. Automatisch vertalen wordt daarna beschreven. Dan beschrijven we database publicatie als vertaalhulpmiddel en als laatste zullen we enkele overige hulpmiddelen beschrijven.

### 5.1 Tekstverwerker

De tekstverwerker is één van de meest gebruikte (elektronische) hulpmiddelen bij het vertalen. Alle teksten wordt elektronisch aangeleverd en de vertaling moet in elektronische vorm teruggestuurd worden.

Met de tekstverwerker moet het in elk geval mogelijk zijn om de vertaling in te voeren, te bewerken, in te lezen en op te slaan. Wil de tekstverwerker goed in het vertaaltraject passen dan moeten aan alle functies eisen gesteld worden.

#### 5.1.1 Invoeren en bewerken van tekst

Vrijwel alle talen gebruiken speciale karakters, bijvoorbeeld letters met accenten en structureel andere karakters als in Japans. De tekstverwerker moet hier goed mee om kunnen gaan. Hoe makkelijk dit is, hangt af van de gebruikte taal.

Tijdens het vertalen van een tekst is het gemakkelijk het origineel en de vertaling tegelijkertijd op het scherm te zien. Het gebruiksgemak is voor een groot gedeelte afhankelijk van de interface van de tekstverwerker. De minst gebruiksvriendelijke interface is wanneer de originele tekst op papier apart gehouden moet worden en alleen de vertaling op het scherm zichtbaar is. De meest gebruiksvriendelijke versie is wanneer het origineel én de vertaling tegelijkertijd op het scherm zichtbaar zijn, terwijl wordt bijgehouden welke zin in het origineel

bij welke zin in de vertaling hoort. Bij verplaatsing in het origineel wordt ook in de vertaling verplaatst en vice versa. Hierdoor zijn altijd de originele zin en de vertaling ervan tegelijkertijd zichtbaar.

### 5.1.2 Inlezen en opslaan van tekst

Tekstverwerkers kunnen meestal bestanden in verschillende bestandsformaten inlezen en soms ook opslaan. Hierbij doen zich twee soorten problemen voor:

1. Niet alle bestandsformaten worden goed ondersteund,
2. Er kunnen speciale karakters in een document voorkomen die niet goed verwerkt worden.

Om dit soort problemen op te lossen kan het beste gekozen worden voor open standaard bestandsformaten. Dit zijn bestandsformaten die gemakkelijk overdraagbaar zijn tussen verschillende programma's. Voorbeelden hiervan zijn SGML, ANSI en Unicode.

## 5.2 Termenbank

Normaal is de definitie van een term: een woord dat in een bepaald vakgebied een bepaalde (andere) betekenis krijgt of alleen in dat vakgebied gebruikt wordt. In de context van termenbanken die gebruikt worden om termen te vertalen wordt deze definitie uitgebreid met: een term is een woord waaraan bepaalde restricties worden gesteld met betrekking tot de betekenis of mogelijke vertalingen.

Een termenbank is een lijst van termen met bijbehorende vertaling. Deze woordenlijst kan in verschillende vormen zijn:

**Op papier** Dit komt overeen met een speciaal soort woordenboek, een lijst van woorden met daarbij de vertaling. Nadeel van deze vorm is dat elke term met de hand opgezocht moet worden en dat bepaalde termen in de tekst misschien verkeerd vertaald worden omdat de vertaler niet weet dat het een term is.

**Elektronisch** Bij de elektronische versie van de termenbank wordt het woord "termenbank" voor de database met termen gebruikt, maar ook voor de interface naar de database. Een elektronische termenbank heeft voordelen ten opzichte van een op papier, hoewel de interface van de termenbank zeer belangrijk is. Hierin zijn twee mogelijkheden:

**Stand-alone** De elektronische termenbank is een apart programma, dat naast de tekstverwerker gestart moet worden wanneer een woord opgezocht moet worden. Dit heeft als nadeel dat de vertaler elke keer van het ene programma, de tekstverwerker, naar het andere programma, de termenbank, moet gaan. Verder moet de vertaler zelf nog beslissen of een woord een term is, zodat hij nog steeds moet beslissen of het woord opgezocht moet worden.

**Plug-in** De elektronische termenbank is onderdeel van de tekstverwerker, zodat elke keer als de vertaler een term tegenkomt die in de termenbank aanwezig is, automatisch de vertaling van de term aangegeven

wordt. Hiermee wordt een zeer consistente vertaling van termen behaald en de termenbank is zeer gemakkelijk in het gebruik.

Het opzoeken van termen in een termenbank kost tijd. Hoewel bij de elektronische versies het opzoeken van de termen veel sneller gaat dan bij de versie op papier, is het van belang dat het opzoeken van de termen niet te veel tijd kost, omdat dat ten koste gaat van de gebruiksvriendelijkheid van de termenbank.

### 5.3 Vertaalgeheugen

De term vertaalgeheugen wordt vaak voor twee verschillende dingen gebruikt. Ten eerste is een vertaalgeheugen een database waarin zinnen en hun vertaling worden opgeslagen. Deze vorm van vertaalgeheugen zullen we "*vertaalgeheugen-database*" noemen. Ten tweede is er het programma dat met behulp van de database probeert zo veel mogelijk tekst in een document te vertalen. Deze vorm van vertaalgeheugen zullen we "*vertaalgeheugen*" blijven noemen.

Een vertaalgeheugen is een database met een interface naar de gebruiker. Het vertaalgeheugen geeft aan wanneer de vertaler een tekstdeel (meestal een zin) gaat vertalen dat reeds in de vertaalgeheugen-database is opgeslagen. Hierdoor hoeft de vertaler de zin niet te vertalen; de vertaalde zin kan direct uit de vertaalgeheugen-database gehaald worden.

Helaas komt het vaak voor dat te vertalen zinnen niet precies in de vertaalgeheugen-database aanwezig zijn, maar dat de vertaalgeheugen-database wel zinnen bevat die erg lijken op de te vertalen zinnen. In dit geval geeft het vertaalgeheugen aan dat de te vertalen zin erg lijkt op een zin in de vertaalgeheugen-database. Te vertalen zinnen die precies in de vertaalgeheugen-database voorkomen worden "*100% match*" genoemd. Zinnen die erg lijken op zinnen in de vertaalgeheugen-database krijgen een lagere score, bijvoorbeeld 80%, afhankelijk van hoe goed de zinnen op elkaar lijken. Dit soort zinnen worden "*fuzzy match*" genoemd.

Een vertaalgeheugen kan op twee verschillende manieren gebruikt worden.

**Voorvertaling** Een document wordt aangeboden aan het vertaalgeheugen.

Het vertaalgeheugen probeert zoveel mogelijk van het document te vertalen. Hierdoor ontstaat een document waarvan gedeelten vertaald zijn. Bij het voorvertalen van een document worden meestal alleen 100% matches gebruikt, hoewel soms een iets minder passende zin ook wordt goedgekeurd. Welke zinnen goed worden gekeurd is in te stellen, maar over het algemeen worden alleen zinnen toegelaten met bijvoorbeeld 90% of hoger. Bij lagere matches worden verkeerde vertalingen geïntroduceerd. Hoe groot het vertaalde gedeelte is, is afhankelijk van het soort document en hoe goed de vertaalgeheugen-database past bij het document.

Bij het genereren van een voorvertaling is niet iemand nodig die kennis heeft van de taal. Het is een volledig automatisch traject. Na de voorvertaling wordt de rest van het document vertaald door een vertaler. Deze kijkt ook de voorvertaalde delen van het document na.

**Interactief** Een vertaler gebruikt het vertaalgeheugen tijdens het vertalen van het document. Als de vertaler een zin gaat vertalen die in de vertaalgeheugen-database is, zal het vertaalgeheugen aanduiden dat die zin (of

een fuzzy match) in de database aanwezig is, waarna de vertaler eventueel de nodige veranderingen aan de zin kan aanbrengen. Hierbij wordt het vertaalgeheugen dus beter gebruikt, omdat het vertaalgeheugen ook voor slechtere fuzzy matches aangezet kan worden. De vertaler moet dan iets meer veranderen in de zin als hij er voor kiest uit te gaan van de aangeboden zin.

Bij het interactief gebruik van een vertaalgeheugen is altijd een vertaler nodig. Hij moet keuzes maken en aangeboden zinnen zo aanpassen dat de vertaling goed is. De vertaler wordt tijdens het vertalen steeds mogelijke vertalingen aangeboden waaruit hij kan kiezen. Dit is ook meestal de manier waarop een vertaalgeheugen gebruikt wordt. De vertaler is degene die de beste keuzes kan maken tussen de aangeboden zinnen. Hierdoor kan er een beter gebruik gemaakt worden van het vertaalgeheugen. In plaats van alleen zeer goed passende zinnen (met een hoge match) kunnen ook wat minder goed passende zinnen gebruikt worden.

Het bijwerken van de vertaalgeheugen database gebeurt over het algemeen aan het eind van een project of na een duidelijk afgebakende stap in het totale project. In elk geval zijn originele documenten vertaald en gecontroleerd. Tijdens het toevoegen wordt de relatie tussen de originele zinnen en de vertaalde zinnen gelegd en de informatie wordt aan de vertaalgeheugen database toegevoegd.

Overigens is het meestal niet zo dat alle vertaalde zinnen worden toegevoegd aan de database. Het vertaalgeheugen bepaald hoe groot de kans is dat een bepaalde zin vaker voorkomt en op basis daarvan wordt een zin wel of niet toegevoegd.

## 5.4 Documentanalyse

Voordat een vertaler een tekst gaat vertalen, wordt meestal eerst een documentanalyse gemaakt. Met behulp van de resultaten van de documentanalyse kan beslist worden welke hulpmiddelen gebruikt worden bij het vertaaltraject.

Bij de documentanalyse wordt vaak het volgende gemeten:

- Het totaal aantal zinnen dat meerdere malen in het document voorkomt. Dit kan door de computer geteld worden.
- Het aantal 100% matches (zinnen die precies zo in de vertaalgeheugen-database voorkomen). Dit is afhankelijk van de vertaalgeheugen-database waarmee gecontroleerd wordt.
- Het aantal zinnen dat niet precies in de vertaalgeheugen-database voorkomt (bijvoorbeeld geen match, 50%-75%, 76%-85%, 86%-95% en 95%-99% matches).

Verder wordt onderzocht:

- Hoe vaak nieuwe versies van het te vertalen document zullen komen.
- Wat voor soort document vertaald moet worden.

Met deze informatie is het mogelijk een analyse te maken waarmee beslist kan worden welke hulpmiddelen gebruikt moeten worden. Het gaat hierbij vooral om het mogelijke gebruik van een vertaalgeheugen.

Zinnen die meerdere malen in het document voorkomen kunnen geëxporteerd worden, waarna de zinnen slechts één keer vertaald hoeven te worden, aan de vertaalgeheugen-database toegevoegd kunnen worden en daarna “automatisch” vertaald kunnen worden.

Als er vaak nieuwe versies van het document zullen komen, is het al snel nuttig om een vertaalgeheugen te gebruiken. Alleen de veranderde delen van het document moeten dan vertaald worden, omdat de reeds vertaalde delen van de tekst uit de vertaalgeheugen-database overgenomen kunnen worden.

Niet elk soort document is geschikt om vertaald te worden met behulp van een vertaalgeheugen. Teksten als handleidingen zijn zeer geschikt, ze bevatten vaak standaard zinnen en veel terminologie. Correspondentie daarentegen bevat zeer weinig standaard zinnen en is dus minder geschikt voor vertaling met behulp van een vertaalgeheugen.

## 5.5 Alignment

Wanneer geen vertaalgeheugen-database beschikbaar is, maar wel er zijn verschillende teksten met vertalingen aanwezig, is het mogelijk een vertaalgeheugen te maken uit de reeds vertaalde teksten. Dit gebeurt met een alignmenthulpmiddel.

Een alignmenthulpmiddel vergelijkt twee documenten, het origineel en de vertaling en deelt de teksten op in tekstdelen; dat zijn over het algemeen zinnen. Deze tekstdelen heten segmenten. Nu is het mogelijk de segmenten van het origineel en in de vertaling aan elkaar gelijk te stellen. De vertaling van de segmenten in het origineel is dan bekend. Deze segmenten worden opgeslagen in een bestand dat als vertaalgeheugen-database gebruikt kan worden.

Hoewel dit traject gemakkelijk lijkt, zijn er toch enkele problemen. Ten eerste moeten de segmenten goed “geknipt” worden. Problemen zijn hier bijvoorbeeld punten in afkortingen, deze zijn immers geen zinseinden, dus geen segmenteinden. Ten tweede moeten de segmenten in het origineel verbonden worden met dezelfde, maar vertaalde segmenten in de vertaling. Problemen komen bijvoorbeeld voor wanneer één zin in het origineel vertaald wordt naar twee zinnen, of andersom.

Hoewel een alignmenthulpmiddel soms fouten introduceert in een vertaalgeheugen-database, kan het een handig hulpmiddel zijn. Met een alignmenthulpmiddel kunnen gemakkelijk uit al bestaande vertalingen vertaalgeheugens gemaakt worden.

## 5.6 Vertaalomgeving

Vaak worden de tot nu toe beschreven hulpmiddelen in een compleet pakket samengebracht. Hierdoor is het gemakkelijk van het ene hulpmiddel naar het andere te gaan. De vertaalomgeving kan over het algemeen goed aangepast worden aan de wensen van de vertaler.

De vertaalomgeving bestaat in elk geval uit een tekstverwerker, maar over het algemeen is ook de termenbank een onderdeel. Daarnaast bevat de vertaalomgeving meestal een vertaalgeheugen met de hulpmiddelen die daarbij horen zoals een tekstanalyse hulpmiddel en een alignmmenthulpmiddel.

Het grote voordeel van een vertaalomgeving is dat de vertaler meer hulp krijgt van de hulpmiddelen. Terwijl de vertaler aan het vertalen is, geeft de termenbank vertalingen van de termen in de zin en het vertaalgeheugen geeft mogelijke vertalingen. Ook kunnen vertaalgeheugens gemakkelijk gegenereerd worden; het analyse hulpmiddel kan het alignmmenthulpmiddel aansturen, dat een direct bruikbare vertaalgeheugen-database genereert.

De vertaalomgeving zorgt ervoor dat de verschillende hulpmiddelen beter op elkaar afgestemd kunnen worden, zodat de vertaler meer hulp krijgt dan bij de hulpmiddelen afzonderlijk.

## 5.7 Automatisch vertalen

Bij automatisch vertalen probeert de computer alle tekst die aangeleverd wordt te vertalen. Hierbij wordt geen gebruik gemaakt van een vertaalgeheugen, maar de tekst wordt geanalyseerd en vanuit de analyse wordt een vertaling geproduceerd.

Als eerste zal een korte beschrijving van het verleden gegeven worden. Daarna wordt een overzicht gegeven van de verschillende soorten automatische vertaal-systemen.

### 5.7.1 Verleden

In de jaren '50 startte een grootscheeps onderzoek naar hoge snelheid en hoge kwaliteit vertalingen van willekeurige teksten door de computer. Dit onderzoek werd vooral gefinancierd door militaire organisaties. Men was over het algemeen zeer optimistisch en dacht binnen afzienbare tijd de bestaande problemen op te lossen en dus een algemeen vertaalsysteem te kunnen ontwikkelen.

In de jaren '60 werd langzaam duidelijk dat de problemen bij het automatisch vertalen groter waren dan gedacht en dat steeds nieuwe problemen gevonden werden. In 1966 publiceerde de "National Academy of Sciences" het ALPAC rapport. Het ALPAC rapport gaf een negatief advies op basis van vier argumenten:

1. De kosten van automatische vertaling waren te hoog in vergelijking tot menselijke vertaling.
2. Er was weinig noodzaak tot automatisch vertalen: er was immers weinig vraag en er waren voldoende vertalers.
3. De kwaliteit van automatisch geproduceerde vertalingen was slecht.
4. De toekomst voorspelde weinig verbetering; er moest eerst maar eens flink wat taalkundig onderzoek worden verricht, speciaal gericht op automatische verwerking van natuurlijke taal, alvorens de ontwikkeling van vertaalsystemen ter hand genomen werd.

Na de publicatie van het ALPAC rapport werd bijna al het onderzoek naar automatisch vertalen gestopt of op een zeer laag pitje gezet.

Begin jaren '80 herleefde de interesse naar het vertalen met de computer. Dit had verschillende oorzaken. Ten eerste waren de computers in de tussentijd een stuk sneller geworden, ze hadden grotere geheugens en betere randapparatuur en ze waren goedkoper geworden. Ten tweede was de vraag naar vertalingen naar aanleiding van de toename van internationale contacten toegenomen. Ten derde had men ingezien dat volledig automatisch vertalen met hoge snelheid en hoge kwaliteit in de nabije toekomst (en mogelijk in de verre toekomst) niet mogelijk was, dus men koos voor een andere aanpak. In plaats van volledig automatische vertaalsystemen werd onderzoek gedaan naar vertaalhulpmiddelen, die het werk van vertalers niet overneemt, maar wel vergemakkelijkt.

Tegenwoordig zijn verschillende vertaalhulpmiddelen commercieel beschikbaar. Dit zijn vooral vertaalgeheugens eventueel gecombineerd met een termenbank. Automatische vertaalsystemen zijn ook beschikbaar, maar deze zijn slechts in een zeer beperkt aantal situaties inzetbaar, omdat dit soort systemen alleen goed werkt met een beperkte deelverzameling van een taal. Dit wil zeggen dat strenge restricties moeten worden gesteld aan de grammatica en het lexicon. Daarnaast moet vaak het document nog nabewerkt worden, waarbij een vertaler de vertaalde tekst nog controleert en de slecht vertaalde stukken tekst herschrijft.

### 5.7.2 Verschillende systemen

Dit soort vertaalsystemen zijn onder te verdelen in drie soorten.

**Directe systemen** Bij een direct systeem wordt de omzetting van brontaal<sup>1</sup>-zin  $x$  naar doeltaal<sup>2</sup>-zin  $y$  uitgevoerd met behulp van één enkel algoritme, dat  $x$  als invoer heeft en  $y$  als uitvoer.

**Transfer systemen** Een transfer systeem bestaat in feite uit drie verschillende algoritmen.

- Een *analyse* algoritme dat brontaal-zin  $x$  herleidt tot een abstractere representatie  $x'$ .
- Een *transfer* algoritme dat de abstracte representatie  $x'$  omzet in een equivalente en eveneens abstracte representatie  $y'$ .
- Een *synthese* algoritme dat met  $y'$  als invoer de gezochte vertaling  $y$  construeert.

**Interlingua systemen** Het idee achter een interlingua systeem is om de analyse van doeltaal-zin  $x$  zo ver door te voeren, dat de resulterende abstractie  $x'$  rechtstreeks geschikt is als invoer van het synthese algoritme dat doeltaal-zin  $y$  dient voort te brengen. Vanwege de spil-functie van  $x'$  in de totale trajectgang wordt dit type systeem ook wel als pivot-systeem aangeduid.

In figuur 5.1 is een voorbeeld gegeven van de verschillende stappen binnen een transfer vertaalsysteem.

<sup>1</sup>De brontaal is de taal van het originele, te vertalen document.

<sup>2</sup>De doeltaal is de taal waarnaar het document vertaald moet worden.

---

**Figuur 5.1** voorbeeld van transfer vertaalsysteem
 

---

Gisteren plukten de kinderen bloemen.

↓analyse

```
#S[SIMPL, DECL]:
(NP[MRV, DEF], SUBJ:
(ART[BEP], DET);
(N[‘kind’], HEAD))
(VP[TRANS], PRED:
(V[‘pluk-3’], PAST], MAINVERB);
(NP[MRV, INDEF], OBJ:
(ART[-], DET);
(N[‘bloem-2’], HEAD)))
(MODP[SIMPL], TIME:
(ADV[‘gisteren’], \_))#
```

↓transfer

```
#S[SIMPL, DECL]:
(NP[MRV, DEF], SUBJ:
(ART[BEP], DET);
(N[‘child’], HEAD))
(VP[TRANS], PRED:
(V[‘pick-5’], CONTIN-PAST], MAINVERB);
(NP[MRV, INDEF], OBJ:
(ART[-], DET);
(N[‘flower’], HEAD)))
(MODP[SIMPL], TIME:
(ADV[‘yesterday’], \_))#
```

↓synthese

The children were picking flowers yesterday.

---



Tegenwoordig zijn enkele van dit soort vertaalsystemen operationeel. De resultaten zijn vrij goed, maar daarbij moet wel opgemerkt worden dat de doeltaal een zeer beperkt gebied bestrijkt en verder worden de invoer teksten voorbewerkt zodat slechts een beperkt aantal grammaticale constructies voorkomt. Voor een vertaalsysteem dat algemene teksten moet kunnen vertalen zijn dit soort systemen voorlopig geen goede oplossing.

## 5.8 Automatische vertaling als zijeffect van database publicatie

Het is mogelijk database publicatie te zien als vertaalmiddel. Zoals beschreven in sectie 3.2 is het mogelijk documenten zo samen te stellen dat gemakkelijk een vertaling van het document ontstaat. Hoewel deze werkwijze erg lijkt op een vertaalgeheugen, er worden immers tekstdelen opgeslagen in een database, zijn hier toch enkele verschillen.

Database publicatie	Vertaalgeheugen
Tekstdelen worden opgeslagen in een database. Wat precies tekstdelen zijn wordt aangegeven door de gebruiker. Over het algemeen zijn dit grotere stukken tekst, bijvoorbeeld alinea's. De gebruiker kan direct bij de inhoud van de database.	Tekstdelen worden opgeslagen in een (interne) database. Het vertaalgeheugen geeft aan wat precies een tekstdeel is en gebruiker kan niet direct bij de inhoud van de database. Over het algemeen werken vertaalgeheugens met tekstdelen op zinsniveau.
De database bevat de vertalingen van de tekstdelen. Alleen de precieze vertalingen worden gebruikt als het document vertaald wordt. Bij database publicatie wordt geen gebruik gemaakt van fuzzy match.	De database bevat vertalingen van de tekstdelen, maar het vertaalgeheugen werkt met fuzzy match. Hierdoor kunnen ook (niet geheel passende) delen van de tekstdelen gebruikt worden.
Bij database publicatie wordt het document samengesteld van delen uit de database. Hierdoor ontstaan "automatisch" vertalingen.	Het vertaalgeheugen wordt gebruikt bij het vertalen van nieuwe teksten (die al bekende delen bevatten).
De gebruiker bepaald welke delen van het document aan de database worden toegevoegd.	Het vertaalgeheugen bepaalt welke zinnen met vertalingen welke zinnen met vertalingen worden toegevoegd aan de database. Dit hoeven dus niet alle vertaalde zinnen te zijn.

## 5.9 Overige hulpmiddelen

In deze sectie worden hulpmiddelen beschreven die misschien niet direct vertaalhulpmiddelen zijn, maar daar wel direct mee te maken hebben. Deze hulpmiddelen helpen niet bij het vertalen. Ze helpen bijvoorbeeld bij het maken van bepaalde keuzes, die met het vertalen te maken hebben of hebben een ondersteunende functie. In de appendix wordt nog een hulpmiddel beschreven dat kan

helpen bij het maken van een keuze tussen verschillende vertaalhulpmiddelen.

### 5.9.1 Versiebeheer programmatuur

Bij een publicatietraject gecombineerd met een vertaaltraject wordt gewerkt met veel verschillende versies van een bepaald document. Om goed bij te kunnen houden in welke versies en op welke locaties een bepaald document aanwezig is moet een versiebeheer systeem gebruikt worden.

Bij vertaalbureaus worden vaak simpele versiebeheer systemen gebruikt. Bijvoorbeeld, voor elk project worden drie mappen gebruikt, een voor de originele documenten, een voor de vertaalde documenten die nog gecontroleerd moeten worden en een voor de uiteindelijke documenten. Bij uitgebreidere projecten is dit soort systemen niet toereikend.

Er bestaan ook enkele elektronische versiebeheer programma's, maar deze zijn in de praktijk niet van voldoende kwaliteit. Ze zijn complex in gebruik en vaak ook niet afgestemd op vertaalprojecten.

### 5.9.2 Evaluatie van vertalingen

Soms is het gewenst een evaluatie van de vertaalde documenten uit te voeren. Bijvoorbeeld wanneer de relatie vanuit het publicatietraject naar het vertaalbureau niet goed is en waarbij getwijfeld wordt aan de kwaliteit van de vertalingen. Een evaluatie van de vertaling kan dan uitkomst bieden. Een andere manier waarop evaluaties van vertalingen hulp kunnen bieden is bij het maken van afspraken in een contract. Met behulp van evaluaties is de kwaliteit van een vertaling meetbaar geworden. In het contract kunnen dus afspraken gemaakt worden over de kwaliteit van de vertaling die meetbaar zijn.

Het probleem bij het meten van de kwaliteit van vertalingen is, dat er niet een algemeen geaccepteerd evaluatiemodel voor vertalingen bestaat. Vragen als "Wat is een goede vertaling?" zijn zeer moeilijk te beantwoorden.

Het is mogelijk uit te gaan van een zelfde soort opzet als bij de evaluatie van vertaalhulpmiddelen, zoals beschreven in appendix A. Er worden bepaalde eisen gesteld waaraan een vertaling moet voldoen, in de vorm van attributen met gewenste waarden. Bij de evaluatie worden de waarden van de attributen gemeten met behulp van gekozen metrieken en kan een evaluatierapport geproduceerd worden.

De problemen bij een aanpak als deze ligt uiteraard bij het kiezen van de attributen en de metrieken. Aan een vertaling kunnen bepaalde eisen gesteld worden, maar vaak zijn die niet of zeer moeilijk meetbaar. Hierbij valt te denken aan de eis dat een document niet meer dan 20 spellingsfouten mag bevatten. De eis is gemakkelijk te stellen, maar het aantal spelfouten in een tekst meten is een stuk moeilijker, zeker als dit automatisch moet gebeuren.

In de praktijk vindt vaak geen evaluatie plaats. Alleen bij grotere bedrijven die internationale vestigingen hebben laten hun vertaalde documenten controleren door mensen in een land waar die taal gesproken wordt.

Op het moment wordt enig onderzoek gedaan naar de evaluatie van vertaalprogrammatuur, zie hiervoor appendix A. Er wordt echter zeer weinig onderzoek gedaan naar het evalueren van vertalingen.

## Hoofdstuk 6

# Algemene problemen bij het vertalen

Zoals in hoofdstuk 3.2 al is aangegeven ontstaan bij het vertalen van informatie enkele specifieke problemen. Deze problemen hebben voor een groot gedeelte te maken met de opmaak van de vertaalde informatie.

In dit hoofdstuk wordt eerst gekeken naar problemen die te maken hebben met karaktersets. Daarna worden problemen met de opmaak van tekst en tabellen beschouwd. Vervolgens worden problemen met het vertalen van teksten in ongebruikelijk bestandsformaat beschreven. Daarna worden beschrijvingen gegeven van problemen met vertaalgeheugens en terminologie. Vervolgens wordt naar de problemen met tags gekeken en als laatste worden problemen met communicatie beschreven.

### 6.1 Problemen met karaktersets

De meeste talen gebruiken een aantal speciale karakters. Hierbij valt te denken aan accenten op letters, zoals é, ø, en ç, maar ook aan totaal andere karakters, zoals gebruikt in Chinees of Arabisch. Dit soort karakters levert een aantal problemen op.

Ten eerste moet de software geschikt zijn voor de gebruikte karakters. De vertaler moet de karakters gemakkelijk in kunnen voeren en bewerken, maar ook moet bijvoorbeeld de software, die gebruikt wordt bij de opmaak, het document met de verschillende karaktersets kunnen bewerken.

Ten tweede moet het document met “vreemde” karakters opgeslagen worden in een uitwisselbaar bestandsformaat, zodat het document ook in andere programma’s in te lezen is. Wanneer de vertaler software gebruikt, die verschillende karaktersets kan verwerken, moet de vertaler het document wel op kunnen slaan in een bestandsformaat dat ook gebruikt kan worden voor de volgende stap in het publicatietraject. Dit hoeft niet een bestandsformaat te zijn dat direct ingelezen kan worden, als het maar te converteren is naar een bestandsformaat dat in te lezen is. Teksten worden aangeleverd in ANSI (met verschillende codepages voor de verschillende karaktersets) en in Unicode.

Ten derde moet er software beschikbaar zijn om kleine wijzigingen aan te kunnen brengen in het document. Wanneer het vertaalde document bijna klaar

is voor productie, maar er moet nog een klein foutje verbeterd worden, is het onhandig om het document eerst weer naar de vertaler te sturen om de fout te verbeteren. Om te zorgen dat dit bij de opdrachtgever kan gebeuren is het nodig software te hebben waarmee kleine veranderingen aangebracht kunnen worden in het vertaalde document. Hiervoor hoeft niet een uitgebreide tekstverwerker gebruikt te worden, een simpele editor volstaat.

## 6.2 Problemen met opmaak van teksten

Als een document vertaald moet worden, krijgt de vertaler een kopie van het originele (opgemaakte) document en een elektronische versie die gebruikt kan worden bij het vertalen zelf. Wanneer het vertaalde document terug komt bij de uitgever is vaak de (meeste) opmaak verloren gegaan. Dit komt omdat er nog geen opmaak in het document verwerkt was of omdat tijdens het vertalen zelf de opmaak verloren is gegaan. Zelfs als de opmaak nog (in de vorm van tags) in het document aanwezig is, moet de opmaak opnieuw aangepast worden, omdat vertaalde tekst meer of minder plaats in kan nemen dan bij het origineel.

Meestal wordt dit opgelost door de vertaalde tekst “passend” te maken. Dit is mogelijk door bijvoorbeeld een iets kleinere of grotere lettergrootte te kiezen of de marges iets aan te passen. Een andere oplossing is om in het originele document al ruimte open te houden en daardoor rekening te houden met vertalingen die meer plaats innemen.

## 6.3 Problemen met tabellen

Bij het vertalen van documenten met tabellen ontstaan een paar specifieke problemen. De vertaler moet weten hoe een tabel eruit ziet, omdat anders de context van de tekst in de tabel onduidelijk kan zijn. Hiervoor krijgt de vertaler een kopie van het origineel. Een ander probleem is dat de opmaak van de tabel anders wordt dan bij het origineel. De vertaler moet hierbij bij de vertaling zo min mogelijk van het origineel afwijken. Wanneer een vertaald stuk tekst in een tabel veel meer plaats inneemt dan bij het origineel kan het zo zijn dat de tekst niet meer in de tabel past.

De oplossing van dit probleem is precies hetzelfde als bij de problemen met de opmaak van teksten. De grootte van het lettertype en de marges kunnen aangepast worden, maar ook in het origineel kan rekening gehouden worden met vertalingen die meer plaats innemen.

## 6.4 Problemen met ongebruikelijke bestandsformaten

Soms bevat de te vertalen tekst ook informatie in een ongebruikelijk formaat. Hierbij valt te denken aan stukken tekst in illustraties, maar in het geval van elektronische uitgaven aan tekst in animaties en eventueel tekst in geluidsfragmenten. We zullen ons hier concentreren op teksten in illustraties.

Teksten in illustraties worden met de hand uit de illustratie gehaald en opgeslagen in een database samen met de coördinaten van de tekst in de illustratie.

Deze database wordt dan (samen met het origineel van de illustratie) opgestuurd naar de vertaler die de teksten vertaalt. Daarna worden de teksten automatisch weer teruggezet in de illustratie. De coördinaten geven aan waar de vertaalde tekst in de illustratie gezet moet worden.

Een betere manier om dit probleem op te lossen is door simpelweg geen tekst in de illustraties toe te laten. Om nu toch te kunnen verwijzen naar delen van een illustratie moet een andere manier gebruikt worden zodat referenties mogelijk zijn. Een mogelijke oplossing is door bijvoorbeeld letters of cijfers aan de illustratie toe te voegen. Daardoor is het mogelijk in de tekst aan delen van de illustratie te refereren.

## 6.5 Problemen met vertaalgeheugens

Een vertaalgeheugen kan een goed hulpmiddel bij het vertalen zijn, waarmee grote delen van documenten automatisch vertaald kunnen worden. Dit is alleen mogelijk als de vertaalgeheugen-database aan bepaalde eisen voldoet:

- *De vertaalgeheugen-database moet goede vertalingen bevatten.* Elke zin in de vertaalgeheugen-database dat met een verkeerde vertaling is verbonden, zal (wanneer de zin niet verbeterd wordt) een verkeerde zin in de uiteindelijke vertaling opleveren.
- *De vertaalgeheugen-database moet goed afgebakende documentdelen bevatten.* Als de vertaalgeheugen-database bijvoorbeeld segmenten bevat die niet goed afgebakend zijn, zullen dat soort segmenten niet passen op stukken tekst van het document. Hierdoor zullen de segmenten niet als vertaling aangeboden worden.
- *De vertaalgeheugen-database moet zo goed mogelijk ingedeeld zijn.* Zinnen die zeer waarschijnlijk nooit meer voorkomen in een stuk tekst, hoeven niet in de vertaalgeheugen-database, terwijl zeer vaak herhalende zinnen juist wel in de vertaalgeheugen-database opgenomen moeten worden. Het indelen van de vertaalgeheugen-database kan dus op twee manieren verkeerd gaan.
  - *De vertaalgeheugen-database bevat teveel zinnen.* Hierdoor wordt het vertaalgeheugen langzaam. Hoewel waarschijnlijk meer zinnen automatisch vertaald kunnen worden, wordt het systeem te langzaam om goed mee te kunnen werken.
  - *De vertaalgeheugen-database bevat te weinig of verkeerde zinnen.* Als de vertaalgeheugen-database te weinig zinnen bevat, worden weinig vertalingen aangeboden tijdens het vertalen. Als de vertaalgeheugen-database verkeerde zinnen, bijvoorbeeld zinnen die waarschijnlijk niet in een ander document voorkomen, bevat, zal het vertaalgeheugen weinig vertalingen aanbieden tijdens het vertalen en tegelijkertijd onnodig langzaam worden, omdat de vertaalgeheugen-database teveel niet nuttige zinnen bevat.
- *De vertaalgeheugen-database moet niet teveel dubbele zinnen bevatten.* Bij sommige vertaalgeheugens is dit niet mogelijk, maar hier moet bij bepaalde

vertaalgeheugens wel rekening mee gehouden worden. De dubbele zinnen vullen de vertaalgeheugen-database zonder dat er nieuwe informatie bij wordt gevoegd.

- *De vertaalgeheugen-database moet goed opgemaakte zinnen bevatten.* Als de vertaalgeheugen-database bijvoorbeeld zinnen bevat met dubbele spaties of extra leestekens, zal een dergelijke zin niet precies passen op dezelfde zin zonder de dubbele spaties of extra leestekens. Het vertaalgeheugen zal dan geen 100% match geven.

## 6.6 Problemen met terminologie

Als een tekst veel terminologie bevat, is het wenselijk de termen consequent hetzelfde te vertalen. Hiervoor wordt meestal een termenlijst aangelegd samen met de vertaling van de termen. Zo een termenlijst kan op papier of elektronisch opgeslagen worden.

Wanneer de termenlijst op papier bij de vertaler aangeleverd wordt, moet de vertaler bij elk woord waarbij hij denkt dat het een term is in de termenlijst opzoeken hoe de term vertaald moet worden. Dit brengt extra werk met zich mee, terwijl ook termen overgeslagen kunnen worden.

Als de termenlijst elektronisch, in een termenbank, is opgeslagen zijn deze problemen er niet. De termenbank kan de vertaler er op wijzen dat een bepaald woord een term is dat in de termenbank is opgeslagen.

Hoewel een elektronische versie van de termenbank niet meer de problemen heeft van de versie op papier, kunnen bij een elektronische versie toch enkele problemen zich voordoen.

- *De termen in de termenbank moeten goed gespeld zijn.* Als een te vertalen term niet goed gespeld is in de termenbank, zal de termenbank niet duidelijk maken dat het een term is. Hierdoor is het mogelijk dat de term niet goed vertaald wordt.
- *De termen in de termenbank moeten up-to-date zijn.* Meestal is dit een iets ander probleem. Wanneer de terminologie binnen een document verandert of anders vertaald moet worden, wordt dit meestal wel in de termenbank veranderd, maar niet in de vertaalgeheugen-database. Zinnen die het vertaalgeheugen vertaalt, bevat andere terminologie dan de termenbank aan zou geven. Dit is alleen op te lossen door, bij veranderende terminologie, niet alleen de termenbank aan te passen, maar ook alle voorkomens van de veranderde termen in de vertaalgeheugen-database.

## 6.7 Problemen met tags

Tags in te vertalen documenten kunnen problemen veroorzaken, maar ook problemen oplossen. Of tags problemen veroorzaken of oplossen is afhankelijk van hoe er met de tags om wordt gegaan.

Als een vertaler tijdens het vertalen tags verwijdert of verandert, kan na het vertalen het document niet of moeilijk verder verwerkt worden. De tags in het document worden gebruikt om bijvoorbeeld het document te voorzien van

opmaakinformatie. Als tags niet in het document verwerkt zijn zoals verwacht wordt, kan het omzetten van de tags naar opmaakinformatie niet automatisch meer gebeuren. Het is een vaak moeilijk en tijdrovend traject om de tags weer op de goede plaats terug te krijgen in het document.

Een oplossing voor dit probleem is te zorgen dat de vertaler de tags niet kan veranderen of verwijderen. Als het programma waarin de vertaler zijn teksten vertaalt, bijvoorbeeld Transit of Trados Translator's Workbench, zorgt dat de vertaler de tags niet kan veranderen of verwijderen zullen de tags binnen het document intact blijven.

Programma's die gebruikt worden bij het invoeren van vertalingen en tegelijkertijd tags kunnen "begrijpen", kunnen verschillende problemen oplossen en leveren extra gebruiksgemak. Door het gebruik van tags is het mogelijk voorvertaalde documenten te bewerken waarbij de originele tekst tegelijkertijd met de vertaling op het scherm staat, terwijl beide of alleen het origineel niet bewerkbaar zijn. Hierdoor is strikter gebruik van het vertaalgeheugen mogelijk, maar is het voor de vertaler wel mogelijk (voorvertaalde) context van de originele zinnen te lezen. Ook is het mogelijk bijvoorbeeld termen en namen voor te vertalen en met tags aanduiden dat deze niet bewerkbaar zijn. Tegelijkertijd kan een betere schatting gemaakt worden van de kosten van de vertaling. Voorvertaalde tekstdelen hoeven alleen gecontroleerd te worden.

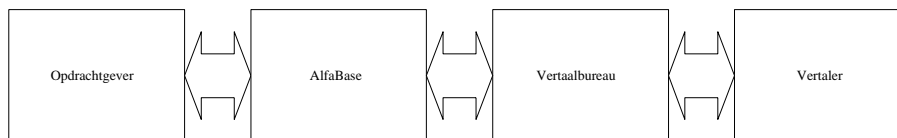
Sommige programma's die nu commercieel beschikbaar zijn herkennen tags in bestanden, maar de meeste programma's laten de vertaler vrij in het toevoegen en verwijderen van die tags. Hierdoor kan de structuur van het document verloren gaan. Het is wenselijk dat alle tags op een zelfde manier in het origineel en in de vertaling van het document staan.

## 6.8 Problemen met communicatie

Door de tussenkomst van een bedrijf dat het publicatie- en vertaaltraject coördineert kunnen problemen ontstaan wat betreft de communicatie. Dit komt voornamelijk door de vele stappen die nodig zijn voordat een vraag of opmerking van bijvoorbeeld vertaler tot de opdrachtgever doorgegeven is.

Het totale traject ziet eruit als weergegeven in figuur 6.1.

**Figuur 6.1** communicatie in publicatie- en vertaaltraject



Van tevoren moet aandacht besteed worden aan de communicatie, omdat bijvoorbeeld een probleem bij de vertaler drie communicatiestappen nodig heeft om bij de opdrachtgever aan te komen. Er moet afgesproken worden hoe de communicatie gecontroleerd wordt. Bij elke stap kan namelijk informatie verloren gaan.





## Hoofdstuk 7

# Interface tussen vertaal- en publicatietraject

Wanneer het vertaal- en publicatietraject worden samengevoegd tot een traject, ontstaan een aantal problemen op de overgang van het ene traject naar het andere. Die problemen hebben te maken met communicatie en met het feit dat de twee trajecten elkaar gedeeltelijk overlappen.

Als eerste worden hier de problemen besproken die te maken hebben met de communicatie tussen de twee trajecten. Daarna worden problemen besproken die te maken hebben met de locatie van verschillende hulpmiddelen.

### 7.1 Communicatie

Bij de combinatie van het vertaal- en publicatietraject werken twee bedrijven samen. Alfa▼Base verzorgt het publicatietraject en een vertaalbureau handelt het vertaaltraject af. Daarbij moet op een bepaald moment informatie van het ene naar het andere bedrijf worden doorgegeven. Het doorgeven gebeurt twee kanten op.

1. Documenten worden ter vertaling van Alfa▼Base naar het vertaalbureau doorgegeven.
2. Vertaalde documenten worden teruggegeven van het vertaalbureau naar Alfa▼Base.

Het is van groot belang dat het heen en weer sturen van documenten goed geregeld is. Het versturen van documenten kan grote problemen opleveren als er geen goede afspraken gemaakt zijn. Hoe goed beide bedrijven ook hun eigen traject voor elkaar hebben, wanneer het overgeven van de documenten niet goed gaat zal het totale traject problemen opleveren.

Om de overgang van het ene bedrijf naar het andere zo goed mogelijk te laten verlopen, zijn goede afspraken nodig. Het is dan ook van cruciaal belang dat er een goede relatie is tussen Alfa▼Base en het vertaalbureau. Het is waarschijnlijk wenselijk dat gekozen wordt voor één vertaalbureau waarmee een duurzame relatie wordt aangegaan. Hierdoor ontstaat een relatie waarbij beide bedrijven

elkaar kunnen vertrouwen en wanneer er problemen ontstaan deze samen op te lossen.

Als er een goede relatie is tussen de twee bedrijven kunnen ook de verschillende trajecten zo goed mogelijk op elkaar afgestemd worden. Hiertoe moeten afspraken gemaakt worden over

- het bestandsformaat van de documenten,
- in hoeverre de documenten vertaald zijn,
- waar welke vertaalhulpmiddelen aanwezig zijn en welke gebruikt gaan worden,
- wie precies wat doet in de verschillende trajecten,
- hoe aan bepaalde kwaliteitsnormen wordt voldaan en hoe dat gemeten wordt,
- voor welke prijs wat gedaan wordt.

De meeste punten zijn reeds besproken in vorige hoofdstukken. Veel punten zijn afhankelijk van de afspraken die gemaakt worden en van de aard van het project. Er is niet één beste oplossing voor de interface tussen Alfa▼Base en het vertaalbureau.

## 7.2 Locatie

Bij het vertalen van een document moeten verschillende stappen doorlopen worden. Van een aantal van deze stappen is niet direct duidelijk wie ze uit moet voeren. Het meest belangrijke is waarschijnlijk dat mensen werk uit handen moeten kunnen geven.

Hier onder worden twee overwegingen gemaakt over de locatie van de hulpmiddelen. Dit moet in principe voor elk hulpmiddel plaatsvinden. Er wordt hier slechts een globale beschrijving gegeven, omdat de specifieke keuzes afhangen van de afwegingen van het specifieke project.

### 7.2.1 Analyse van het document

Zoals het vertaaltraject nu beschreven is, kan het analyseren van het te vertalen document op verschillende locaties plaatsvinden. Elke keer vindt een vergelijkbare analyse plaats. Met de analyse wordt bekeken welke vertaalhulpmiddelen gebruikt kunnen worden en wat de prijs wordt van de vertaling.

Aangezien bij de analyse een aantal parameters zijn in te stellen, is het mogelijk de analyse een bepaalde richting op te sturen, zodat juist veel of juist weinig gebruik gemaakt zou kunnen worden van vertaalhulpmiddelen.

- *Analyse van het document bij de opdrachtgever.* De opdrachtgever is over het algemeen iemand die niet de expertise heeft om een goede analyse uit te voeren. Verder hebben opdrachtgevers vaak te optimistische ideeën over het mogelijke gebruik van vertaalhulpmiddelen. Deze ideeën ontstaan ook door gebrek aan ervaring met problemen bij het (geautomatiseerde) vertaaltraject. De opdrachtgever zal de uitkomst van de analyse zo laten

worden dat zeer grote delen van de tekst met behulp van vertaalhulpmiddelen vertaald kunnen worden. De gedeelten die vertaald worden met behulp van vertaalmiddelen zijn namelijk goedkoper dan de gedeelten die vertaald worden zonder vertaalhulpmiddelen.

- *Analyse van het document bij Alfa▼Base.* Alfa▼Base zal meestal een betere kennis hebben van het analysehulpmiddel, dus Alfa▼Base kan waarschijnlijk een betere analyse maken dan de opdrachtgever. Alfa▼Base zal echter misschien toch een analyse maken waaruit blijkt dat (te) veel delen van de tekst met vertaalhulpmiddelen vertaald kan worden.
- *Analyse van het document bij de vertaler.* De vertaler heeft voordelen bij een analyse waaruit blijkt dat zeer weinig gebruik gemaakt kan worden van vertaalhulpmiddelen. Het heeft als voordeel dat de vertaler meer betaald krijgt, met de hand vertalen is duurder dan met behulp van vertaalhulpmiddelen. Ook is het vaak zo dat de vertaler niet volledig tevreden is met de vertaalhulpmiddelen die hij moet gebruiken, alhoewel dit minder het geval is wanneer een vertaler langer met een bepaald vertaalhulpmiddel werkt.
- *Analyse van het document bij het vertaalbureau.* Het vertaalbureau zit precies tussen deze twee uitersten in. Aan de ene kant wil de opdrachtgever dat de vertaling zo goedkoop mogelijk is, dus veel gebruik van vertaalhulpmiddelen, maar aan de andere kant wil de vertaler zo veel mogelijk geld voor de vertaling, dus zo min mogelijk gebruik van vertaalhulpmiddelen. Het vertaalbureau moet proberen hier een balans te vinden, waarbij de kosten en het gebruik van de vertaalhulpmiddelen realistisch zijn. Wanneer de opdrachtgever en de vertaler goede relaties hebben met het vertaalbureau zullen ze beiden inzien wanneer het vertaalbureau een realistische analyse maakt. Het vertaalbureau heeft waarschijnlijk meer ervaring in het maken van analyses dan de opdrachtgever en kan goed schatten in hoeverre een vertaalhulpmiddel echt bruikbaar is bij het vertaaltraject.

### 7.2.2 Vertalen van het document met een vertaalgeheugen

Als er na de analyse van het document gekozen wordt om een vertaalgeheugen te gebruiken, ontstaat bijna dezelfde situatie. Het is mogelijk een vertaalgeheugen “stand-alone” te gebruiken, waarbij het document zoveel mogelijk automatisch vertaald wordt en het is mogelijk een vertaalgeheugen “on line” te gebruiken, waarbij tijdens het vertalen aan wordt gegeven dat een bepaalde zin in de vertaalgeheugen-database aanwezig is. Het voorvertalen van het document kan dus weer op meerdere plaatsen gebeuren, namelijk bij de opdrachtgever, bij het vertaalbureau en bij de vertaler.

- *Vertalen met een vertaalgeheugen bij de opdrachtgever.* Ervan uitgaande dat er een vertaalgeheugen-database van tevoren aanwezig is, is het mogelijk dat de opdrachtgever een voorvertaling maakt van het te vertalen document. Na de voorvertaling levert hij het voorvertaalde document aan het vertaalbureau of de vertaler. Deze aanpak heeft voor- en nadelen. Voordelen zijn dat de opdrachtgever de vertaalgeheugen-database in eigen bezit houdt en dat hij ziet hoe groot het reeds vertaalde gedeelte is in het

te vertalen document. Het nadeel is dat als de vertaalgeheugen-database van slechte kwaliteit is, de vertaler erg veel werk heeft aan het verbeteren van de reeds vertaalde delen van het document.

- *Vertalen met een vertaalgeheugen bij Alfa▼Base.* Alfa▼Base zal net als de opdrachtgever een voorvertaling kunnen maken. Alfa▼Base heeft zelf geen vertalers in dienst die het vertaalgeheugen on line kunnen gebruiken. Alfa▼Base heeft vaak wel een betere kennis van het vertaalgeheugen dan de opdrachtgever, dus Alfa▼Base zou beter de voorvertaling kunnen maken. Hiervoor moet de opdrachtgever uiteraard wel de vertaalgeheugen-database uit handen geven.
- *Vertalen met een vertaalgeheugen bij het vertaalbureau.* Aangezien het vertaalbureau meer een coördinerende werking heeft en weinig te maken heeft met het verwerken van de inhoud van documenten is dit niet de meest praktische oplossing. Verder heeft deze werkwijze hetzelfde nadeel als bij het voorvertalen bij de opdrachtgever, maar hierbij moet de opdrachtgever de vertaalgeheugen-database ook nog uit handen geven.
- *Vertalen met een vertaalgeheugen bij de vertaler.* Als de vertaler van een document een goede vertaalgeheugen-database tot zijn beschikking heeft en ermee om weet te gaan, kan dat een goede hulp zijn bij het vertalen van het document. De meeste vertaalgeheugens kunnen ook interactief gebruikt worden, zodat de vertaler tijdens het vertalen aanwijzingen krijgt. Als er toch fouten in de vertaalgeheugen-database aanwezig zijn, kan de vertaler kiezen de vertaling uit de vertaalgeheugen-database niet te gebruiken of de vertaling uit de vertaalgeheugen-database aan te passen. Hierdoor is het ook mogelijk zinnen uit de vertaalgeheugen-database te gebruiken die geen 100% match hebben, maar iets minder goed passen. Dit lijkt dan ook in de meeste gevallen de beste oplossing.

## Hoofdstuk 8

# Aanbeveling voor verbeteringen

Het traject zoals beschreven in sectie 3.4 is ongeveer hoe het op dit moment werkt bij Alfa▼*Base*. Dit is in het algemeen de ideale opstelling voor een dergelijk traject. Hoewel elk project anders is, kan er van die opstelling uitgegaan worden en de precieze parameters kunnen naderhand bijgesteld worden. Mogelijkheden hierbij zijn bijvoorbeeld of er gebruik wordt gemaakt van vertaalhulpmiddelen en welke, of er gebruik wordt gemaakt van database publicatie en of dit invloed heeft op het vertaaltraject.

Hieronder zullen een aantal punten aangestipt worden die belangrijk zijn in het algemene traject, onafhankelijk van de precieze instelling. Als eerste wordt de relatie met een vertaalbureau behandeld, waarbij de nadruk ligt op afspraken over de keuze van vertaalhulpmiddelen. Daarna worden verbeteringen aan de interne organisatie beschreven. Als laatste wordt het belang van afspraken met de opdrachtgever besproken.

### 8.1 Relatie met een vertaalbureau

Zoals al eerder is aangegeven is de relatie tussen Alfa▼*Base* en een vertaalbureau van groot belang. Bij een goede relatie kan gemakkelijker overleg plaats vinden. Hierdoor kunnen beide bedrijven aangeven wat voor hen het beste werkt, waardoor geen onnodige eisen worden gesteld. Beide bedrijven kunnen informatie geven over datgene waar zij het meest van weten.

#### 8.1.1 Vertaalhulpmiddelen

Een van de problemen die zich voor zullen doen heeft te maken met de keuze van de vertaalhulpmiddelen. Vooral bij grotere projecten zal de kans zeer groot zijn dat vertaalhulpmiddelen gebruikt zullen worden en afhankelijk van de ervaring met de verschillende vertaalhulpmiddelen zal er een keuze gemaakt moeten worden.

Over het algemeen moet de keuze aan het vertaalbureau zijn, omdat het vertaalbureau en hun vertalers de hulpmiddelen moeten gebruiken. Toch kan Alfa▼*Base* ook enkele eisen stellen aan de vertaalhulpmiddelen. Deze eisen

hebben dan vooral te maken met de eventuele conversie van documenten na vertaling. Daarbij mag uiteraard zo min mogelijk informatie verloren gaan.

Wanneer de keuze van de vertaalhulpmiddelen gemaakt is, komt de vraag wáár wélke vertaalhulpmiddelen gebruikt gaan worden. Voor een aantal vertaalhulpmiddelen is dit duidelijk.

**Tekstverwerker** Een tekstverwerker is bij de vertalers aanwezig, omdat de vertalers uiteraard de vertaling in moeten kunnen voeren. Ook bij *Alfa▼Base* moet een tekstverwerker aanwezig zijn, dit om kleine veranderingen in het document door te kunnen voeren zonder dat het document terug naar de vertaler moet.

**Termenbank** De termenbank kan op verschillende plaatsen staan, afhankelijk van de gemaakte afspraken. Het is mogelijk dat, voordat het document naar het vertaalbureau gestuurd wordt, alle terminologie vertaald wordt. In dit geval zou de termenbank bij *Alfa▼Base* of de opdrachtgever moeten staan, die de terminologie in het document vertaalt. Een andere mogelijkheid is dat de termenbank bij het vertaalbureau staat. In dit geval verzorgt het vertaalbureau de vertaling van de terminologie. Een derde mogelijkheid is dat de termenbank bij de vertaler staat. De vertaler gebruikt de termenbank tijdens het vertalen. In dit geval wordt de terminologie niet van tevoren vertaald, maar tijdens het vertaaltraject. In de praktijk zal de plaats waar de termenbank aanwezig is per project kunnen verschillen. Het is natuurlijk niet zo dat de termenbank maar op één plaats kan staan; meerdere locaties tegelijkertijd is ook mogelijk. Wanneer de termenbanken uitgebreid worden wat betreft netwerk functionaliteit, is het ook mogelijk dat de termenbank op één locatie aanwezig is, maar dat iedereen er gebruik van kan maken via een netwerk.

**Vertaalgeheugen** Voor het vertaalgeheugen gelden dezelfde mogelijkheden als bij de termenbank. Het vertaalgeheugen kan gebruikt worden bij het voorvertalen van het document. Dit kan gebeuren bij *Alfa▼Base*, de opdrachtgever of het vertaalbureau, maar het kan ook gebruikt worden tijdens het vertalen zelf. Aangezien het vertaalbureau meer een logistieke functie heeft, blijft over de mogelijkheid van het voorvertalen of vertalen met behulp van het vertaalgeheugen. Afhankelijk van de afspraken kan hier een keuze gemaakt worden. Het vertaalgeheugen naast het (voor-)vertalen, ook gebruikt om een analyse te maken van het te vertalen document. Hoewel de analyse meestal bij het vertaalbureau plaats vindt, zie ook het volgende punt, wil de opdrachtgever of *Alfa▼Base* de analyse kunnen controleren. Daarvoor moet een vertaalgeheugen aanwezig zijn. Over het algemeen zullen dus op verschillende plaatsen kopieën van het vertaalgeheugen nodig zijn. Ook hier geldt dat, wanneer de vertaalgeheugens betere netwerk functionaliteit krijgt, het vertaalgeheugen op één plaats fysiek aanwezig is, maar door meerdere mensen (op verschillende locaties) gebruikt kan worden via een netwerk.

**Documentanalyse** De documentanalyse is op verschillende plaatsen mogelijk. Ook hier is het van belang wat de gemaakte afspraken zijn. Over het algemeen heeft documentanalyse bij het vertaalbureau de meeste voordelen. Het vertaalbureau heeft de benodigde kennis. Het vertaalbureau heeft een

overzicht van de te vertalen documenten en moet zorgen dat de relaties naar Alfa▼Base en de vertalers goed blijven. Alfa▼Base en de vertaler maken meestal hun eigen analyses om de analyse van het vertaalbureau te kunnen controleren.

**Alignment** Alignment van documenten moet gebeuren in samenwerking tussen Alfa▼Base en het vertaalbureau of zelfs ook met vertalers. Dit is afhankelijk van hoe goed de alignmenthulpmiddelen zijn en hoe goed gestructureerd de vertaalde documenten zijn. Over het algemeen is dit een eenmalige stap.

**Vertaalomgeving** Vertaalomgevingen worden speciaal voor vertalers ontwikkeld. De meest logische plaats voor een vertaalomgeving is dus bij de vertalers. Toch kan ook de vertaalomgeving bij het vertaalbureau en zelfs bij Alfa▼Base of de opdrachtgever handig zijn, afhankelijk van de geïntegreerde hulpmiddelen in de vertaalomgeving.

**Automatisch vertalen** Wanneer gebruik gemaakt gaat worden van automatisch vertalen, is dit om de vertaalstap zo klein en zo goedkoop mogelijk te maken. In dit geval zal het automatisch vertalen dus bij Alfa▼Base plaats vinden. Dit is een vrij theoretische opzet, omdat het automatisch vertalen van vrije tekst voorlopig niet de benodigde kwaliteit oplevert en het vertalen van gelimiteerde tekst niet of weinig gebruikt wordt.

**Database publicatie** Net als bij automatisch vertalen is dit een techniek die gebruikt wordt om het vertaaltraject zoveel mogelijk uit te schakelen. Met behulp van database publicatie kunnen automatisch volledig vertaalde teksten gegenereerd worden, die alleen gecontroleerd hoeven te worden door vertalers. Het mag duidelijk zijn dat, net als bij automatisch vertalen, database publicatie hulpmiddelen bij Alfa▼Base of bij de opdrachtgever aanwezig zullen zijn, niet bij een vertaalbureau of de vertaler.

**Evaluatie van vertalingen** Het evalueren van vertalingen kan op verschillende plaatsen gebeuren. De vertaler kan hiermee controleren in hoeverre hij aan de eisen van het vertaalbureau of de opdrachtgever voldoet. Het vertaalbureau kan controleren of de vertaler een goede vertaling geproduceerd heeft en of deze vertaling voldoet aan de gestelde eisen. Alfa▼Base kan controleren of de vertaling voldoet aan de eisen die met het vertaalbureau zijn afgesproken en of de vertaling voldoet aan de eisen die de opdrachtgever gesteld heeft.

Over het algemeen heeft elk hulpmiddel zijn eigen functie die afhankelijk is van de locatie. Sommige hulpmiddelen zijn zeer specifiek, sommige zijn meer algemeen. Waar precies welk hulpmiddel gebruikt moet worden is zeer afhankelijk van de gemaakte afspraken, de precieze invulling van het project en van de relatie tussen de verschillende deelnemers.

## 8.2 Interne organisatie

Het publiceren in meerdere talen brengt ook een verandering in de interne organisatie te weeg. Wanneer bij een eentalig publicatietraject één document

voor elke stap een beperkt aantal versies heeft, wordt dit bij meerdere talen verveelvoudigd. Dit kan bij een redelijk aantal talen een explosieve groei van de hoeveelheid van versies van documenten opleveren.

Met deze hoeveelheid van documenten is een goed beheer nodig. Het documentbeheer kan aangeven op welke locatie een bepaald document is. Daarnaast is het mogelijk de status van het document te weten te komen. Daarbij kan gedacht worden aan “nog niet vertaald”, “vertaald” of “ter correctie”.

Een beheersysteem moet echter ook verschillende versies van documenten kunnen bijhouden. Wanneer het traject gestart is met een bepaald document, maar naderhand wordt een verbeterde versie van het document aangeleverd, moet duidelijk zijn wat er kan gebeuren. Over het algemeen is het nodig het hele publicatietraject opnieuw te starten met de nieuwe versie van het document.

### 8.3 Afspraken met opdrachtgevers

Elk project verschilt van andere projecten, dus het is van essentieel belang goede afspraken te maken met de opdrachtgevers. Naast de afspraken die bij een eentalig publicatietraject gemaakt worden, moeten bij een meertalig publicatietraject een aantal extra afspraken gemaakt worden.

- Er moet worden afgesproken in welke talen gepubliceerd moet worden. Daarbij moet nagegaan worden of het vertaalbureau ook in deze talen kan vertalen.
- De opdrachtgever moet aan kunnen geven wat voor kwaliteit hij voor de vertalingen wil hebben. Gezien de opdrachtgevers van *Alfa▼Base* zullen deze over het algemeen kwalitatief hoogstaande vertalingen moeten zijn. Hierover moeten uiteraard ook afspraken gemaakt zijn met het vertaalbureau. Deze afspraken betreffen de kwaliteit van de vertalingen, maar ook de manier waarop de kwaliteit van de vertalingen gecontroleerd gaat worden.
- Er moet worden aangegeven in hoeverre de opdrachtgever het totale traject zal uitbesteden. Hoewel *Alfa▼Base* een volledig meertalig publicatietraject kan aanbieden kan de opdrachtgever kiezen niet het volledige, maar een gedeelte van het publicatietraject te gebruiken.



# Hoofdstuk 9

## Case

In dit hoofdstuk zal een voorbeeld behandeld worden. Dit is een aangepast voorbeeld uit de praktijk. Het voorbeeld is aangepast zodat een beter overzicht gegeven kan worden van het totale traject. Bij een echt voorbeeld uit de praktijk ontstaan vaak specifieke problemen die een globale beschrijving zoals hier gewenst is onmogelijk maken. Het probleem is zo aangepast dat verschillende problemen en oplossingen hier zichtbaar worden.

Hier wordt een mogelijke oplossing voor het probleem beschreven. Er zijn uiteraard meerdere oplossingen mogelijk, onder andere afhankelijk van verschillende wensen van de opdrachtgever en eigenschappen van de documentatie. Omdat dit een simpele beschrijving is van een project zullen we proberen de belangrijkste punten te beschrijven. In de praktijk komen altijd nog andere problemen naar boven. Dit soort problemen zal hier niet beschreven worden om de beschrijving van het project niet onduidelijker te maken.

Als eerste zal een probleemstelling gegeven worden. Daarna wordt de organisatie en workflow van het complete traject beschreven. Vervolgens zullen de gebruikte technieken en hulpmiddelen worden beschreven. Als laatste wordt een overzicht van de kosten gegeven.

### 9.1 Probleemstelling

In deze sectie zal eerst de huidige situatie worden besproken, gevolgd door een beschrijving van veranderingen in het volledige proces en de verwachtingen daarvan.

#### 9.1.1 Huidige situatie

In deze sectie wordt beschreven hoe het publicatietraject bij het fictieve softwarebedrijf PinSoft georganiseerd is. Het publicatietraject wordt gebruikt bij het publiceren van handleidingen en verpakkingen. Dit is dan ook het traject waar hier op geconcentreerd wordt.

#### Beschrijving van het bedrijf

Het bedrijf PinSoft maakt en verkoopt computerspellen. De nadruk daarbij ligt op flipperkasten voor de computer. De bestaande spellen naast de computerflip-

perkasten zullen in de toekomst niet meer door PinSoft ondersteund worden. Er wordt hier daarom vanuit gegaan dat de flipperkasten de enige producten zijn. Het assortiment bestaat uit vier verschillende spellen, die worden uitgegeven in Duitsland en Nederland. Elk jaar verandert het assortiment naar vier andere spellen.

### **Documenten**

Voor elk nieuw spel wordt elke keer helemaal opnieuw een handleiding geschreven. De ontwerpers van de spellen schrijven en verzorgen de opmaak van de handleidingen. Een handleiding wordt geschreven in het Engels. Elke handleiding bestaat uit een beschrijving van de installatie en een beschrijving van het spel zelf. Hoewel elke handleiding in elk geval deze twee onderdelen bevat, is buiten deze onderverdeling geen sprake van een standaard indeling. Ook is geen sprake van een standaard opmaak.

Er wordt altijd slechts één versie van een handleiding geschreven. De handleiding wordt nadat deze gepubliceerd is, niet meer aangepast. Als de handleiding echt grote fouten bevat wordt eventueel op een apart blad de errata vermeld. Deze werkwijze zal ook niet in de toekomst veranderd worden, dus zal hier geen aandacht besteedt worden aan versiebeheer.

De verpakking van de spellen bevat zo min mogelijk tekst. Alleen de titel van het spel en de systeemeisen van de computer zijn aanwezig. De systeemeisen zijn in het Engels. In plaats van een spelbeschrijving op de doos van het spel zijn meerdere schermafdrucken weergegeven.

### **Workflow**

Het schrijven en opmaken van de handleiding wordt bij PinSoft gedaan met behulp van een tekstverwerker. Het opgemaakte document wordt bij een drukkerij aangeleverd. Bij de drukkerij wordt het document gedrukt; er wordt daar niets aan het document veranderd. De gedrukte handleidingen worden bij PinSoft aangeleverd.

De opmaak van de verpakking wordt met een tekenprogramma gemaakt. Het bestand met de lay-out van de verpakking wordt bij dezelfde drukkerij aangeleverd. De verpakking wordt daar gedrukt en in elkaar gezet. Ook de verpakking wordt dan net als de handleiding naar PinSoft teruggestuurd.

PinSoft laat een tweede bedrijf de software op cd-rom of floppydisk zetten. Het spel wordt aan het bedrijf geleverd, dat de software op de juiste informatiedrager zet. Daarna worden de informatiedragers bij PinSoft aangeleverd.

Op een gegeven moment heeft PinSoft dan de gedrukte handleidingen, de verpakkingen en de informatiedragers. PinSoft schakelt dan een laatste bedrijf in, dat de informatiedrager samen met de handleiding in een verpakking plaatst. Dit bedrijf stuurt dan de uiteindelijke pakketten door naar de distributeur.

### **Standaardisatie**

Voor de ontwikkeling van de software is sinds kort een standaard omgeving ontwikkeld, zodat gemakkelijk nieuwe spellen ontworpen kunnen worden. Door deze standaard omgeving is de besturing van de verschillende spellen hetzelfde. Buiten de besturing is elk spel verschillend.

Naast een standaard ontwikkelomgeving is ook een standaard installatieprogramma ontwikkeld. Deze installatieprocedure kan op twee manieren gebruikt worden, afhankelijk van of het spel op een cd-rom of op een floppydisk uitgegeven wordt. De tekst in het installatieprogramma is net als de handleidingen en de verpakking in het Engels.

### 9.1.2 Verwachtingen

PinSoft verwacht dat in de nabije toekomst een aantal zaken zullen veranderen. Omdat het huidige publicatie- en afwerkingstraject omslachtig is, willen ze dit verbeteren. Daarnaast moet het publicatietraject worden uitgebreid met een vertaalstap.

In deze sectie wordt aangegeven wat precies de verwachtingen zijn van de producten in de toekomst en wat verwacht wordt van het vernieuwde publicatie- en afwerkingstraject.

#### Uitbreiding

Door de ontwikkeling van de standaard omgeving kunnen sneller nieuwe spellen ontworpen worden. De verwachting is dan ook dat binnen korte tijd het assortiment wordt uitgebreid tot acht of meer spellen.

Naast de uitbreiding in het aantal spellen wil PinSoft de spellen ook in andere landen uitgeven. Hoewel het Engels een algemeen gangbare taal is in computer-software en handleidingen, wil PinSoft door het vertalen van de handleidingen een betere ondersteuning geven. De handleidingen zullen in het Duits, Engels, Frans en Nederlands beschikbaar moeten komen. In de verre toekomst zou het kunnen dat de handleidingen ook in het Japans beschikbaar moeten komen. Hoewel dit niet direct gebeurt moet het traject wel met deze mogelijkheid rekening houden.

De verpakking zal nog steeds zo min mogelijk tekst bevatten. De tekst op de verpakking zal alleen in het Engels beschikbaar zijn. Het idee is dat de scherm-afdrukken een goed beeld geven van het spel. Een uitgebreide beschrijving is daarom niet nodig.

De tekst in het installatieprogramma zal ook, in plaats van alleen in het Engels, in de verschillende talen worden vertaald. Dit komt de ondersteuning net als het vertalen van de handleidingen ten goede.

#### Organisatie

Vanwege de verwachte groei in aantal spellen en verschillende versies van de handleidingen wil PinSoft iemand aantrekken om de handleidingen te verzorgen. De keuze om iemand anders dan de ontwerpers de handleidingen te laten schrijven heeft twee redenen. Ten eerste hoeven de ontwerpers de handleidingen niet meer te schrijven, zodat ze meer tijd hebben om nieuwe spellen te ontwerpen en implementeren. Ten tweede verwacht PinSoft dat de handleidingen een betere kwaliteit zullen krijgen. De ontwerpers zijn oorspronkelijk aangenomen voor het ontwerpen van de spellen, niet voor het schrijven van een handleiding.

## 9.2 Organisatie en workflow

In deze sectie zullen de veranderingen in de organisatie besproken worden. Als eerste zullen de veranderingen binnen PinSoft behandeld worden. Daarna worden de veranderingen in organisatie buiten PinSoft beschreven. Het gaat daarbij vooral over het afwerkingstraject.

### Interne organisatie

Zoals PinSoft zelf al aangaf wordt één persoon verantwoordelijk voor het schrijven van de handleidingen. PinSoft gaf zelf de eerste twee voordelen al aan, maar er zijn meer voordelen:

- De ontwerpers hoeven de handleidingen niet meer te schrijven, dus zij hebben meer tijd over voor het ontwerpen.
- De kwaliteit van de handleidingen kan beter worden omdat de ontwerpers geen tekstschrijvers zijn.
- De tekstschrijver krijgt een spil-functie. Aan de ene kant heeft hij direct contact met de ontwerpers die hem de nodige informatie geven en aan de andere kant houdt hij overzicht op het publicatietraject.
- Het schrijfproces kan gemakkelijker gestandaardiseerd en geautomatiseerd worden omdat er maar één persoon direct bij betrokken is.

Het standaardiseren van het schrijfproces heeft zelf weer veel voordelen:

- Omdat de handleiding een standaard structuur krijgt, zal een nieuwe handleiding alle nodige informatie bevatten. De standaardisatie werkt als een soort controlelijst; alle punten die de handleiding moeten bevatten staan ook echt in de handleiding.
- Het is beter mogelijk een standaard opmaak toe te passen, zodat de verschillende handleidingen uiterlijk op elkaar lijken. Een standaard opmaak kan meehelpen aan realiseren van een huisstijl.
- Standaardiseren helpt mee aan mogelijk hergebruik van teksten.
- Het standaardiseren maakt het mogelijk de handleidingen om te zetten naar andere media dan papier.

In principe is het belangrijkste dat de documentatie gestandaardiseerd en dus gestructureerd wordt. Het feit dat één persoon de handleidingen gaat produceren is niet essentieel, maar maakt het standaardiseren wel gemakkelijker. In dit geval hoeft slechts één persoon zich aan te passen aan de veranderingen in het proces. Als de ontwerpers de handleidingen bleven schrijven hadden ze zich allemaal moeten aanpassen.

### Externe organisatie

Na het schrijven van de handleiding ontstaat in de huidige situatie een complex traject. Het product wordt in veel verschillende stappen samengesteld en na elke stap komt alle informatie weer terecht bij PinSoft. Dit is een tijdrovend en onnodig complex geheel. Daarbij komt nog dat dit traject nog uitgebreider wordt wanneer een vertaalstap wordt toegevoegd.

Een bedrijf dat de organisatie van dit hele traject overneemt is een groot deel van de oplossing van dit probleem. Een bedrijf als Alfa▼*Base* kan de organisatie van het traject overnemen en tegelijkertijd een deel van het traject zelf uitvoeren.

Alfa▼*Base* neemt het traject over, waardoor PinSoft alleen de handleiding hoeft te schrijven en hoeft te zorgen dat het spel op een informatiedrager geleverd wordt. Hierdoor wordt de organisatie van de afwerking voor PinSoft een stuk gemakkelijker.

Het spel wordt aan een bedrijf geleverd dat het op een informatiedrager zet en aan Alfa▼*Base* levert. Het document wordt aan Alfa▼*Base* geleverd, waar de rest van het traject wordt afgehandeld. Als eerste wordt het document vertaald, dit gebeurt bij een vertaalbureau, maar Alfa▼*Base* organiseert en beheert deze stap. Daarna vindt een volledig publicatietraject plaats. Als laatste worden de handleidingen samen met de informatiedragers in de verpakking gestopt en verstuurd naar de distributeurs.

## 9.3 Techniek

In deze sectie worden de gebruikte technieken en programmatuur beschreven. Als eerste worden de technieken beschreven die gebruikt worden in het meertalig publicatietraject. Daarna wordt de gebruikte programmatuur beschreven.

### 9.3.1 Publicatietraject

Een handleiding van PinSoft is in te delen in twee afzonderlijke delen die beide een iets ander publicatietraject zullen krijgen. Het ene deel is de beschrijving van de installatie procedure die vrijwel geheel hergebruikt kan worden en het andere deel is de spelhandleiding die voor elk spel opnieuw geschreven wordt. Beide delen zullen hier afzonderlijk worden beschreven.

#### Beschrijving van de installatie procedure

PinSoft heeft een standaard installatieprocedure ontworpen. De installatie zal dus voor elk spel vrijwel hetzelfde zijn. Dit levert een groot voordeel op bij het schrijven van dit deel van de handleiding. Zeer grote delen van de beschrijving kunnen hergebruikt worden.

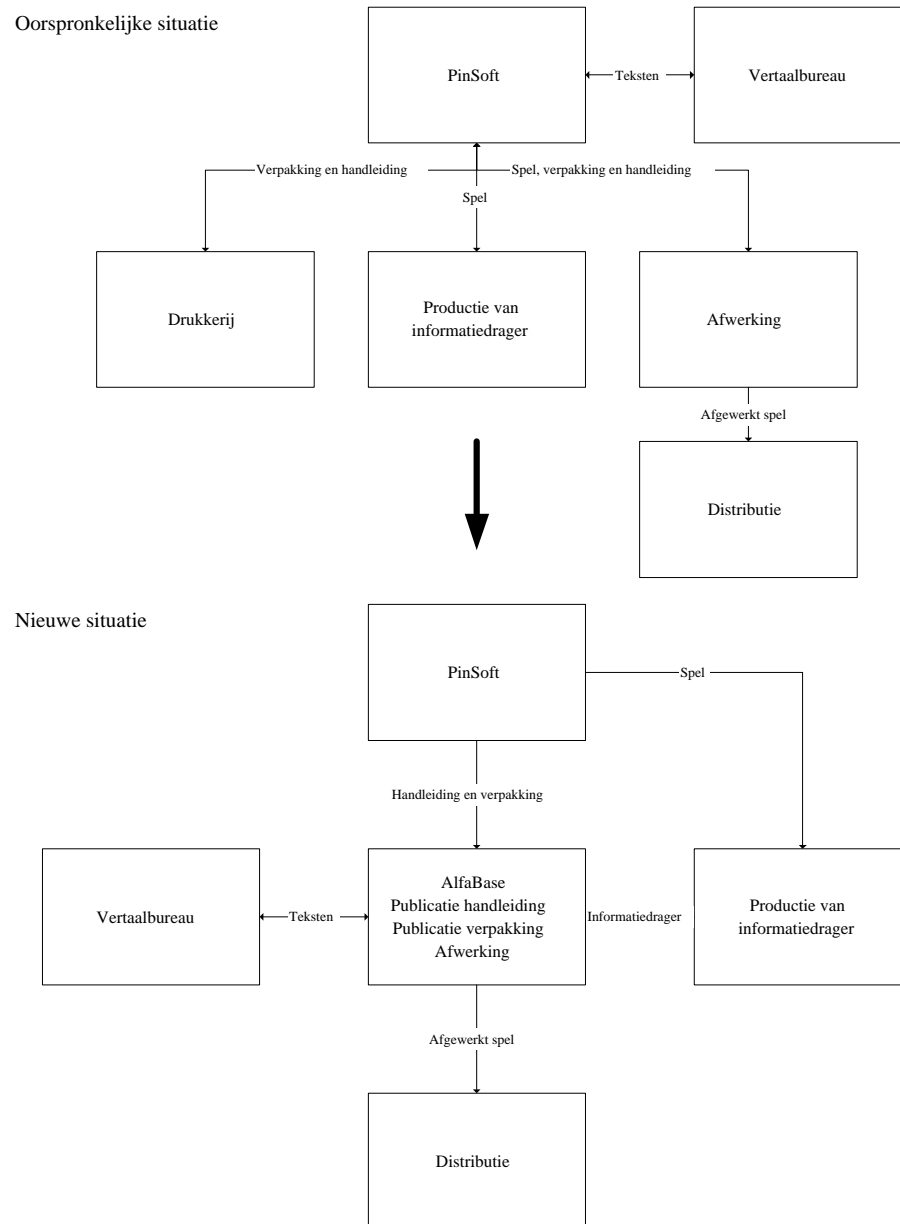
Er zijn twee instanties van de installatie procedure: één om van cd-rom te installeren en één om van floppydisk te installeren. Daarnaast zijn er enkele spelspecifieke instellingen, zoals de directory waarin geïnstalleerd wordt.

Voor beide instanties wordt eenmalig een handleiding in verschillende talen geschreven. Elke volgende keer kan een van deze instanties hergebruikt worden, waarbij alleen de spelspecifieke instellingen aangepast moeten worden. Wanneer is aangegeven in het origineel waar de spelspecifieke instellingen in het document

---

**Figuur 9.1** organisatie van PinSoft
 

---



staan, kan het deel van de handleiding zonder tussenkomst van een vertaler in meerdere talen geproduceerd worden.

Het aangeven van specifieke instellingen gebeurt met behulp van SGML tags. De spelspecifieke informatie kan dan gevonden worden door te zoeken op bepaalde SGML tags, bijvoorbeeld “<SPECIFIC>”. De tekst tussen de SGML tags moet dan per spel veranderd worden, maar hier hoeft geen taalafhankelijke informatie ingevoegd te worden.

In feite is dit een vergaande vorm van database publicatie. Er hoeft slechts gekozen te worden tussen twee verschillende teksten. Het grote voordeel hiervan is dus dat de teksten slechts eenmalig vertaald hoeven te worden door een vertaler. Na de vertaling kunnen de teksten geheel hergebruikt worden. Dat scheelt vertaalkosten en de tijd die het kost om de tekst te laten vertalen.

Vanwege het kleine aantal documenten (voor elke taal twee) is hier geen database nodig. Wanneer een bepaalde tekst nodig is, wordt deze gekopieerd en ingevoegd in het totale document. Daarbij blijven de documenten altijd volledig; er worden geen tekstdelen geselecteerd zoals in echte database publicatie.

### Spelbeschrijving

De spelbeschrijving is voor elk spel verschillend. Het is dus niet mogelijk om grote delen tekst te hergebruiken zoals bij de beschrijving van de installatie procedure. De spelbeschrijving wordt dus voor elk spel volledig opnieuw geschreven. Er wordt voor gekozen een standaard terminologie te gebruiken. Hierdoor wordt het taalgebruik in de beschrijving consistent en gemakkelijker te vertalen.

Elke spelbeschrijving heeft een gelijksoortige opzet. Elke beschrijving bevat bijvoorbeeld een beschrijving van de besturing van het spel, de specifieke regels en de puntentelling. SGML wordt gebruikt om de spelbeschrijving een standaard structuur te geven. Dit heeft meerdere voordelen:

- Door het structureren van de beschrijving met SGML worden alle belangrijke onderdelen aan de beschrijving toegevoegd. Er worden geen onderdelen vergeten, omdat SGML de structuur van de spelbeschrijving forceert. De spelbeschrijving moet een inleiding, beschrijving van de besturing, enzovoorts bevatten.
- Vanwege de standaard opzet wordt de beschrijving goed leesbaar. Als een uiteindelijke klant meerdere spellen heeft, kan hij gemakkelijk in de documenten zoeken; ze hebben allen dezelfde structuur.
- Door het gebruik van SGML kan automatisch een standaard opmaak gerealiseerd worden, zelfs bij de vertalingen.

De vertalingen van de spelbeschrijving worden gemaakt door het vertaalbureau. Omdat bepaalde zinnen vaak terugkomen in verschillende spelbeschrijvingen wordt gebruik gemaakt van een vertaalgeheugen. Het vertaalgeheugen is aanwezig bij de vertaler zelf, zodat er optimaal gebruik gemaakt kan worden van de vertalingen die aanwezig zijn in het vertaalgeheugen.

#### 9.3.2 Programmatuur

Tijdens het schrijven, vertalen en publiceren worden verschillende programma's gebruikt. In deze sectie worden de gebruikte programmatuur beschreven. Daar-

bij worden de voor- en nadelen van de verschillende programma's genoemd.

Als eerste zal de programmatuur van het schrijfproces beschreven worden. Daarna worden de programma's die gebruikt worden bij het vertaaltraject beschreven. Als laatste zal de programmatuur van het publicatietraject worden beschreven.

### Schrijfproces

Het schrijfproces is ingedeeld in twee verschillende trajecten.

**Beschrijving van de installatieprocedure** Zoals was aangegeven is hier de keuze uit de twee verschillende manieren van installeren. Beide manieren hebben een standaard beschrijving. Deze is eenmalig geschreven en vertaald en kan daarna herhaaldelijk gebruikt worden. Dit is een simpele vorm van database publicatie. Er is de keuze uit twee teksten, die daarna iets aangepast moeten worden.

**Spelbeschrijving** De spelbeschrijving wordt elke keer opnieuw gemaakt. Hierbij is het hergebruiken van tekst minder goed mogelijk. Het zou kunnen dat toch delen van de tekst herbruikbaar zijn. Hierbij valt te denken aan de beschrijving van de besturing van het spel.

De beschrijving van de installatieprocedure kan direct vertaald bij Alfa▼*Base* aangeleverd worden. De spelspecifieke delen van de beschrijving zijn dan al bij de opdrachtgever met een tekstverwerker aangepast. De spelspecifieke delen zijn immers taalonafhankelijk, er is dus geen vertaler voor nodig.

De spelbeschrijving moet elke keer opnieuw geschreven worden. Hierbij wordt gebruik gemaakt van een tekstverwerker. In principe kan dit een simpele tekstverwerker zijn. De enige eis hieraan is dat een tekstbestand in een formaat opgeslagen kan worden zodat het aansluit op de volgende stap in het publicatie- en vertaaltraject.

Vanwege de gewenste structurering van het document wordt als extra eis gesteld dat de tekstverwerker met SGML om kan gaan. Dit kan op drie manieren gerealiseerd worden.

**Een tekstverwerker met SGML uitbreiding** Dit is een standaard tekstverwerker, bijvoorbeeld Word, waarbij een speciale uitbreiding is aangebracht. "Near and far Author", "WordPerfect SGML Edition" en "SGML Author for Word" zijn voorbeelden van zo een dergelijke uitbreiding. De uitbreidingen zijn afzonderlijke producten die naderhand aan een bepaalde tekstverwerker toegevoegd kan worden.

**Een tekstverwerker met SGML kennis ingebouwd** Dit is een tekstverwerker die de kennis van SGML als deel van het programma heeft. Hierdoor is een betere ondersteuning van SGML mogelijk. "Framemaker +SGML" is hier een voorbeeld van.

**Een SGML editor** Dit is een speciaal soort tekstverwerker die speciaal ontwikkeld is om SGML te kunnen verwerken. De nadruk ligt hierbij op het SGML, niet op het tekstverwerken. Bijvoorbeeld "Adept" en "SoftQuad Author/Editor" zijn SGML editors.



Welk soort “tekstverwerker” gekozen wordt hangt af van specifieke eisen en mogelijkheden. De systemen hebben voor- en nadelen. Als eerste zullen de voor- en nadelen van de tekstverwerker met SGML uitbreiding besproken worden, daarna worden de voor- en nadelen van de SGML editor behandeld. De tekstverwerker met SGML kennis ingebouwd wordt hier niet verder besproken omdat nog te weinig programmatuur in deze groep valt. In principe heeft deze groep een combinatie van de voor- en nadelen van de beide andere groepen.

De voordelen van een tekstverwerker met SGML uitbreiding zijn dat de gebruiker gemakkelijk zijn oude documenten kan hergebruiken als de tekstverwerker deze bestanden in kan lezen. Een ander voordeel is dat de gebruiker geen nieuw programma hoeft te leren. De uitbreiding is geïntegreerd in de tekstverwerker, dus komen er alleen een aantal extra functies bij.

Een van de nadelen van een tekstverwerker met SGML uitbreiding is dat het invoeren van SGML in de tekstverwerker meestal vrij complex is. De persoon die een tekst invoert moet vaak bekend zijn met de structuur van de DTD. De DTD beschrijft welke tag op welke plaats mag komen te staan. Een ander nadeel is dat bepaalde functies van de tekstverwerker niet gebruikt mogen worden, omdat die functies geen SGML tags opleveren. Hierbij valt te denken aan een onderlijnd stuk tekst. Dit mag niet met de gewone functie van de tekstverwerker gekozen worden, maar moet met een SGML tag aangegeven worden. Er is in een tekstverwerker met een SGML uitbreiding meestal geen strenge restrictie op dit soort invoerfouten. De invoerfouten leveren slechte of zelfs foute SGML op.

De SGML editor heeft ook voor- en nadelen. Een van de voordelen is dat de SGML editor de auteur strikt aan de DTD houdt. Hierdoor wordt met een SGML editor altijd correct SGML afgeleverd. Een ander voordeel is dat een SGML editor over het algemeen niet zo uitgebreid is als een tekstverwerker, zodat de SGML editor veel gemakkelijker in het gebruik is.

Een nadeel van de SGML editor is dat een nieuw programma geleerd moet worden. In de praktijk blijkt die niet een zeer groot nadeel te zijn. Ten opzichte van de tekstverwerker met SGML uitbreiding is de tijd om bekend te raken met de SGML editor zeer kort.

Hoewel de meeste mensen de overstap naar een SGML editor niet snel durven nemen, is dit vaak wel de beste keuze. De tekstverwerker met SGML uitbreiding levert meestal slecht SGML op. Hierdoor zijn na het invoeren uitgebreide controle- en conversiestappen nodig.

Ook voor dit project is gekozen voor een SGML editor. Omdat een nieuw persoon wordt aangetrokken om de documentatie te verwerken, hoeft maar één persoon te leren met de SGML editor te werken. Niemand hoeft over te stappen van een tekstverwerker naar een SGML editor. Het overstappen van een tekstverwerker naar een SGML editor kan wat problemen opleveren, maar dat is dus in dit geval niet aan de orde. De SGML editor levert altijd correct SGML af. Hierdoor zijn de SGML bestanden goed in de volgende stappen van het publicatietraject te verwerken. Er zijn geen extra conversie- of controlestappen nodig.

### **Vertaaltraject**

Na het schrijven zullen de documenten door een vertaaltraject gaan. Na het vertaaltraject zijn de documenten in verschillende talen beschikbaar. Bij het

vertaaltraject worden ook een aantal hulpmiddelen gebruikt. Deze zouden afgestemd moeten zijn op het schrijfproces en op het publicatietraject.

Bij het vertaaltraject wordt in elk geval gebruik gemaakt van een tekstverwerker. Deze tekstverwerker zou in elk geval de SGML tags in het document moeten houden. Als dit niet het geval is, moet na het vertalen de documenten opnieuw gecontroleerd en in het ergste geval opnieuw van tags voorzien worden.

Naast de tekstverwerker kan in dit geval goed gebruik worden gemaakt van een termenbank en een vertaalgeheugen. De documenten die vertaald moeten worden bevat bepaalde terminologie. De terminologie kan automatisch met een termenbank vertaald worden. Voorbeelden van terminologie zijn “the shift key”, “floppydisk”, maar ook “jet-bumper”, “slingshot”. Verder komen in de documenten bepaalde zinnen meerdere malen voor. Deze kunnen automatisch vertaald worden met behulp van een vertaalgeheugen. Voorbeelden van dit soort zinnen zijn “press enter” en “insert floppydisk”.

Om het vertaaltraject aan te laten sluiten op het schrijfproces en het publicatietraject moet gekozen worden voor een tekstverwerker en een vertaalgeheugen die goed met SGML om kunnen gaan. Ook moeten de hulpmiddelen onderling ook goed op elkaar aansluiten.

Niet alle vertaalgeheugens kunnen goed met SGML overweg. Bij sommige is het zelfs mogelijk tags te verwijderen. Ook heeft niet elk vertaalgeheugen een goede interface met een termenbank. Bij elk verschillend project moet apart gekeken worden welke vertaalgeheugens en termenbanken gebruikt worden. In dit project blijven twee bekende vertaalgeheugens over:

- Trados Translator’s Workbench
- Transit

Beide zijn vertaalgeheugens gecombineerd met een termenbank. Ook kunnen beide pakketten omgaan met SGML tags. Trados kan werken samen met Word of WordPerfect als tekstverwerker, terwijl Transit een eigen tekstverwerker bevat. Transit ondersteunt alle hier gebruikte talen, maar ondersteunt niet het Japans, terwijl Trados dat wel doet. In dit project wordt daarom gekozen voor Trados Translator’s Workbench.

### Publicatietraject

Het publicatietraject kan uitgaan van bestanden in SGML. Deze SGML bestanden kunnen gemakkelijk omgezet worden in bestanden met zet-codes. Met dit soort bestanden is het mogelijk direct de handleidingen te drukken.

Bij het publicatietraject wordt conversieprogrammatuur gebruikt om SGML om te zetten naar opmaakcodes en opmaakprogrammatuur om met de hand de documenten op te maken.

Voor de conversieprogrammatuur kan gekozen worden uit verschillende pakketten. Omdat Alfa▼Base over het algemeen werkt met Omnimark en Perl worden deze pakketten ook gebruikt voor de conversie van de vertaalde documenten. Ook voor opmaakprogrammatuur zijn verschillende pakketten mogelijk. De bekende programmatuur is:

- Quark XPress
- FrameMaker

- XYVision

De verschillende pakketten hebben elk hun eigen toepassingsgebied. Quark XPress wordt vooral gebruikt om kleine documenten op te maken. Grafische documenten zoals brochures en folder materiaal worden hiermee opgemaakt. FrameMaker wordt vaak gebruikt voor grotere documenten. XYVision wordt ook voor grote documenten gebruikt, maar bij XYVision ligt de nadruk op het zo automatisch mogelijk opmaken van de documenten. De laatste twee programma's ontlopen elkaar (in elk geval wat het meertalige traject betreft) zeer weinig. In dit specifieke traject is gekozen voor XYVision vanwege de gemakkelijke conversie van SGML naar XYVision opmaakcodes.

## 9.4 Kosten

In deze sectie worden de kosten van een meertalig publicatietraject behandeld. Er zullen hier geen bedragen genoemd worden, omdat deze in de loop van de tijd kunnen veranderen. Deze sectie bevat een overzicht van de verschillende kosten die gemaakt worden in deze case.

Als eerste zullen de kosten worden besproken die nodig zijn om dit project op te starten. Daarna zullen de productiekosten worden besproken.

### 9.4.1 Opstartkosten

De opstartkosten vallen uiteen in verschillende groepen:

- Onderzoekskosten:
  - Kosten van het onderzoek en beschrijven van het project
  - Kosten van de aanpassingen van het standaardpakket (zoals beschreven in deze scriptie)
- Kosten van hard- en software:
  - Kosten van de benodigde computers en harde schijven
  - (Aanschaf-)kosten van de software
    - \* SGML editor
    - \* Trados Translator's Workbench
    - \* Omnimark/Perl
    - \* XYVision
- Kosten van de initiële opzet van het traject:
  - Kosten van het ontwikkelen van een vertaalgeheugen en termenbank
  - Kosten van inwerkperiode en opleidingen van de schrijver
  - Kosten van het ontwerpen van de DTD

### 9.4.2 Productiekosten

De productiekosten ontstaan op verschillende plaatsen:

- Het schrijfproces:
  - Arbeidsloon van de schrijver
- Het vertaaltraject:
  - Kosten van het vertaalbureau
- Het publicatietraject:
  - Kosten van de conversie
  - Kosten van het opmaken
  - Kosten van het publiceren
  - Kosten van de afwerking
- De productie van de informatiedrager (cd-rom/floppydisk)

# Hoofdstuk 10

## Toekomstvisie

Vooraf in de informatica kunnen nieuwe ontwikkelingen een grote invloed hebben op verschillende gebieden. Snellere computers bijvoorbeeld kunnen applicaties uitvoeren die vroeger niet voor mogelijk werden geacht. Dit soort veranderingen hebben ook invloed op het publicatie- en vertaaltraject.

Hoewel sinds het begin van “vertalen met de computer” er slechts zeer geleidelijk vooruitgang geboekt wordt, is het mogelijk dat door groeiende interesse van het bedrijfsleven en onderzoeksscholen er de komende jaren snel vooruitgang geboekt zal worden. Hoe snel vooruitgang geboekt zal worden is zeer moeilijk te voorspellen, dus wij zullen ons zo min mogelijk bezighouden met een tijdsschatting.

Als eerste zal gekeken worden naar de veranderingen die te verwachten zijn op het gebied van vertaalhulpmiddelen. Daarna worden veranderingen beschreven die een inslag zullen hebben in het vertaal- en het publicatietraject. De beschrijvingen zullen over het algemeen zo realistisch mogelijk zijn.

Op bepaalde plaatsen zal in de tekst een kader geplaatst worden. De tekst binnen de kaders zijn beschrijvingen van een verre toekomst. Uiteraard is zo een beschrijving een persoonlijke visie, waaraan geen vaste waarde gehecht kan worden. Dit soort tekst moet meer gezien worden als een soort “star trek” beschrijving.

### 10.1 Vertaalhulpmiddelen

Er zijn verschillende oorzaken voor veranderingen van vertaalhulpmiddelen in de toekomst. Een van de belangrijkste oorzaken is dat hoewel het ALPAC rapport gezorgd heeft dat een tijd zeer weinig onderzoek naar vertalen met de computer gedaan werd, mensen en bedrijven zich langzaam maar zeker weer op het gebied van vertalen met computers durven te begeven. Dit is uiteraard een voorwaarde voor vernieuwingen in de toekomst. Andere oorzaken zijn beschreven in de volgende secties.

#### 10.1.1 Hardware

Bij het vertaaltraject worden verschillende soorten hardware gebruikt. We zullen hier de belangrijkste soorten beschrijven samen met de verwachting van

de veranderingen in de toekomst.

De computer is de belangrijkste hardware bij een vertaler. De computer zal de komende jaren waarschijnlijk net zoals de afgelopen jaren steeds sneller worden en de prijs zal blijven dalen. Wanneer de vertaler snellere computers met meer intern geheugen ter beschikking krijgt, is het mogelijk complexere programmatuur te gebruiken. Dit is waarschijnlijk de belangrijkste verandering voor de vertaler wat hardware betreft.

Het verwisselbaar opslagmedium zal ook in de toekomst veranderen. Hoewel tot op heden de floppydisk nog het meest gebruikt is, zal binnenkort een vervanger zich aanbieden. Veranderingen hierbij zullen te maken hebben met betrouwbaarheid, opslagcapaciteit en snelheid. Hoewel deze verandering hoog nodig is, zal dit niet tot een grote verandering in de werkwijze van de vertaler leiden.

De vertaler krijgt steeds vaker documenten via de telefoonlijn, met behulp van een modem. Zoals nu al te merken is, worden modems steeds sneller. Hoewel documenten hierdoor sneller naar de vertaler gestuurd kunnen worden, zal ook deze verandering weinig verschil in de werking van het vertaaltraject opleveren.

Naast de bestaande hardware kan het zo zijn dat nieuwe soorten hardware hun intrede doen. Een realistisch voorbeeld hierbij is de *netwerk computer (NC)*. De NC is een simpele computer die alle informatie, de software en andere bestanden, van een netwerk haalt. Omdat de NC zelf eigenlijk geen randapparatuur aangesloten heeft, is het een zeer goedkope computer. Voor de gebruikte software hoeft alleen betaald te worden als het gebruikt wordt. Een programma wordt dus niet gekocht, maar "gehuurd".

Een nieuwe techniek die met hardware en software samen te maken heeft is *distributed systems (DS)*. DS valt in verschillende technieken uiteen. DS kan gebruikt worden bij *distributed computing*, waarbij een gebruiker veel rekencapaciteit kan krijgen door gebruik te maken van verschillende computers tegelijkertijd, via een netwerk. Meerdere computers (*compute servers*) die verbonden zijn door een netwerk verzorgen de rekencapaciteit. DS kan ook gebruikt worden voor *distributed databases*. Hierbij kan gezocht worden in databases die verdeeld zijn over meerdere computers, zonder dat de gebruiker weet dat de database over verschillende computers verdeeld is. Dit heeft als voordeel dat de database erg groot kan worden zonder dat de zoektijd erg verslechtert. De databases worden toegankelijk door een aantal computer aan het netwerk, de *database servers*. Een overzicht van distributed systems is weergegeven in figuur 10.1.

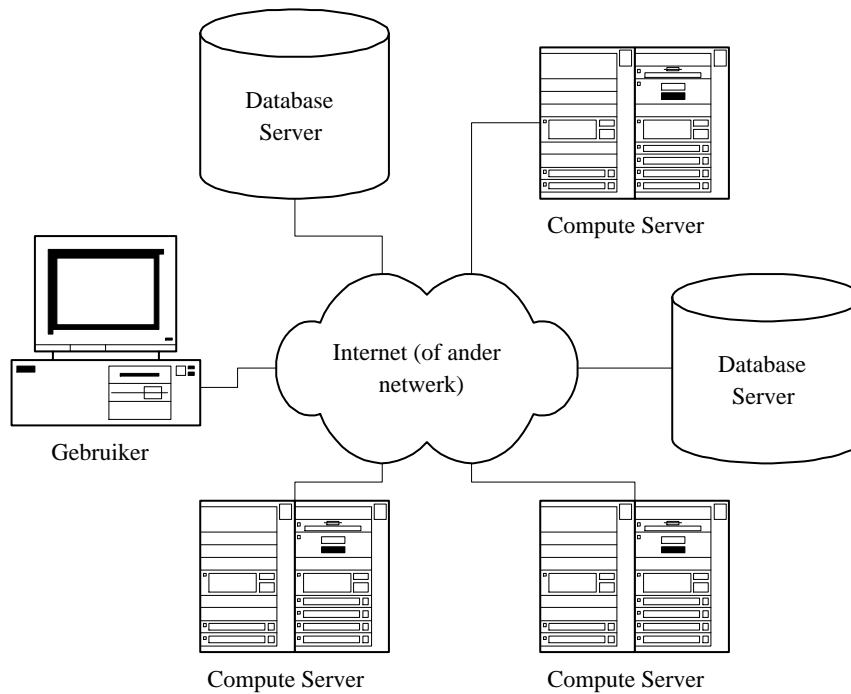
Door deze nieuwe technieken kunnen betere prestaties geleverd worden bij bijvoorbeeld vertaalgeheugens (denk aan distributed databases) en automatisch vertalen (denk aan distributed computing).

<p>Er zal een groot wereldwijd netwerk komen, veel groter dan het huidige internet. Iedereen krijgt toegang tot dat netwerk; een netwerkverbinding wordt net zo gewoon als een telefoonaansluiting. Met dit netwerk komt een zeer grote rekencapaciteit beschikbaar door middel van distributed computing. Voor de gebruiker lijkt het alsof de programmatuur op één computer loopt, maar in werkelijkheid kunnen dit er duizenden zijn, allemaal verbonden door het netwerk.</p>
---

---

**Figuur 10.1** distributed systems
 

---



### 10.1.2 Software

Met het automatiseren van een traject, in dit geval de vertaal- en publicatietrajecten, worden de veel herhalende handelingen overgenomen door de computer. Het automatiseren heeft voor- en nadelen, die eigenlijk bij het automatiseren van elk traject aanwezig zijn. Het geautomatiseerde traject kan sneller verlopen, er zijn minder probleemgevallen binnen het traject, het werk wordt interessanter, enzovoorts. Als nadeel is altijd de vraag in hoeverre deze voordelen echt behaald kunnen worden. Wordt het traject echt sneller doorlopen? Zijn er echt minder probleemgevallen, of zijn er na het automatiseren juist andere probleemgevallen?

In de praktijk blijkt dat het automatiseren van het vertaaltraject wel degelijk een winst oplevert. De vertalingen bevatten een consistentere taalgebruik, termen worden consistentere vertaald en de vertaler hoeft steeds herhalende stukken tekst niet elke keer opnieuw te vertalen. Het automatiseren van het publicatietraject heeft in de praktijk ook veel voordelen. Het traject verloopt soepeler, omdat de afzonderlijke stappen beter op elkaar afgestemd kunnen worden.

Bij het automatiseren wordt gebruik gemaakt van verschillende hulpmiddelen. De gebruikte hulpmiddelen zullen hier beschreven worden. De nadruk zal liggen op vertaalhulpmiddelen.

De afgelopen jaren worden steeds meer vertaalhulpmiddelen (commercieel) beschikbaar gesteld. Verschillende softwarebedrijven zien in dat er vraag is naar vertaalhulpmiddelen. Aan automatisch vertalen wordt (nog) niet veel aandacht besteed.

Waarschijnlijk de belangrijkste trend binnen de vertaalsoftware betreft de veranderingen in de programmatuur. Veranderingen in de software zijn mogelijk in twee richtingen.

1. Er zal meer programmatuur op de markt komen met (ongeveer) de functionaliteit van software die nu beschikbaar is. Hierdoor is er meer keus tussen de software pakketten, waarbij de verschillen tussen de programmatuur zich vooral zullen voordoen bij de gebruikersinterface, benodigde hardware, snelheid en beschikbare talen.
2. Er zal programmatuur met nieuwe functionaliteit beschikbaar komen. Hierbij valt vooral te denken aan ondersteunende software. Voorbeelden hiervan zijn
  - Versiebeheer programmatuur: Hiermee is het mogelijk bij te houden waar verschillende versies van een document zijn en welke status de versies hebben.
  - Meertalige spelling- en grammaticacontrole: Met dit soort programmatuur is het mogelijk de spelling en grammatica van de vertaalde documenten te controleren.
  - Vertaalgeheugen beheerprogrammatuur: Dit soort programmatuur maakt het mogelijk gemakkelijk de database van het vertaalgeheugen aan te passen. Hiermee kunnen fouten verbeterd worden in de database, er kunnen nieuwe teksten met vertalingen toegevoegd worden en er kunnen teksten uitgehaald worden.
  - Programmatuur voor database publicatie: Dit is een compleet pakket met hulpmiddelen. Stukken tekst moeten in een database gezet kunnen worden, maar ook worden verwijderd en veranderd. Verder moet programmatuur aanwezig zijn om een nieuw document te maken. Zo een programma moet een soort tekstverwerker zijn gecombineerd met een goed zoekstelsel om de tekstdelen te vinden.

### Tekstverwerkers

Tekstverwerkers zijn al jaren beschikbaar. In de afgelopen jaren is er eigenlijk weinig veranderd. De interface is gebruiksvriendelijker geworden en de mogelijkheden zijn uitgebreider geworden. De uitgebreidere mogelijkheden zijn zeer specifiek. Veel mensen, ook vertalers, zullen deze mogelijkheden niet gebruiken.

Het belangrijkste van een tekstverwerker is dat tekst ingevoerd en bewerkt kan worden. Tekstverwerkers kunnen echter niet goed met verschillende talen omgaan. Invoeren en bewerken van tekst lijkt redelijk taalafhankelijk, maar toch ontstaan hier een aantal problemen.

Spellingscontrole en eventueel grammaticacontrole is zeker taalafhankelijk. Dit soort controlesystemen zullen in de toekomst beter worden en in meer talen beschikbaar komen. Dit soort systemen kan de vertaler helpen bij het invoeren, controleren en veranderen van vertalingen.

Het gebruik van verschillende karakterset moet gemakkelijker gemaakt worden. Hierbij moet niet gedacht worden aan karaktersets als Times en Courier, maar aan karaktersets waarmee bijvoorbeeld Grieks of Arabisch beschreven kan worden.



De huidige PC's werken intern met *codepages*. Een codepage bevat de karakterset voor een bepaalde taal. De computer kan goed werken met die bepaalde taal, maar wanneer een andere taal gebruikt wordt, moet overgeschakeld worden naar een andere codepage. De huidige generatie besturingssystemen kan nog niet werken met meerdere codepages tegelijkertijd, dus meertalige tekstverwerkers zijn daardoor moeilijk te realiseren.

Met Unicode, een nieuwe standaard, is het mogelijk speciale karakters te selecteren. Unicode zal zeer waarschijnlijk een belangrijke rol gaan spelen in programmatuur waar speciale karakters gebruikt worden. Door het gebruik van Unicode wordt niet alleen een consistent gebruik van lettertypen verkregen, maar ook wordt bestandsuitwisseling van bestanden met "vreemde" lettertypen tussen verschillende programma's mogelijk.

Unicode is een systeem dat lijkt op ASCII, het beschrijft met welk getal een bepaald karakter aangeduid kan worden. Er zijn meerdere verschillen tussen Unicode en ASCII, maar het belangrijkste verschil is dat Unicode alle belangrijke karaktersets die bestaan kan verwerken.

Wanneer veel verschillende karakters in een taal mogelijk zijn, ontstaat een probleem met de interface. Het moet op een makkelijke manier mogelijk zijn verschillende karakters te selecteren. Het invoeren van een karakter mag geen tijdrovende taak zijn.

In de toekomst wordt meer en meer programmatuur gebaseerd op Unicode. Hierdoor is het mogelijk met veel verschillende karaktersets tegelijkertijd te werken. Ook is het mogelijk gemakkelijk bestanden uit te wisselen volgens deze standaard.

Tekstverwerkers zullen meer een soort dicteerapparaat worden. Het is mogelijk de tekstverwerker tekst te laten voorlezen en op te laten nemen door tegen de computer te praten. Uiteraard zullen tekstverwerkers perfect om kunnen gaan met verschillende talen. De tekstverwerkers worden uitgerust met een hulpmiddel waarmee gemakkelijk speciale karakters ingevoerd kunnen worden. Dit kan bijvoorbeeld zijn in de vorm van een toetsenbord waarbij de karakters op het toetsenbord aangepast kunnen worden, zodat op display's op de toetsen de speciale karakters te zien zijn.

## Vertaalgeheugens

Tegenwoordig zijn verschillende goed werkende vertaalgeheugens beschikbaar. De functionaliteit van de vertaalgeheugens ontloopt elkaar erg weinig en de onderliggende zoektechnieken zijn vaak (in elk geval volgens de specificaties) bijna hetzelfde. Toch blijken vertalers nog niet allemaal met vertaalgeheugens te willen werken. Dit heeft verschillende oorzaken.

- *Vertalers denken dat de computers hun werk overnemen.* Het mag duidelijk zijn dat dit niet het geval zal zijn. De kans dat computers (ooit) het complete vertaaltraject zullen overnemen is zeer klein. Daar zijn namelijk automatische vertaalsystemen voor nodig, die met een hoge kwaliteit kunnen vertalen. Dat soort systemen zijn in elk geval in de nabije toekomst niet te verwachten.

- *Het ontwerp van de interface van de vertaalgeheugens kan beter.* Wanneer een vertaler een vertaalgeheugen gebruikt bij het vertalen van een document, moet erg veel informatie tegelijkertijd op het scherm beschikbaar zijn. Het (te vertalen) origineel is op het scherm aanwezig, maar ook de vertaling, aanwijzingen uit het vertaalgeheugen bij fuzzy matches, aanwijzingen uit de termenbank en eventueel een hulpmiddel waarmee “vreemde” karakters ingevoerd kunnen worden. Door al deze informatie verliest de vertaler het overzicht op het beeldscherm.
- *Vertaalgeheugens zouden gemakkelijker in gebruik moeten zijn.* Dit probleem heeft direct te maken met de interface van de vertaalgeheugens. Vooral bij het invoeren of veranderen van teksten die “vreemde” karakters bevatten levert bij vertaalgeheugens problemen op, maar ook bijvoorbeeld het kopiëren van stukken tekst moet gemakkelijk gedaan kunnen worden.
- *De vertaalgeheugens moeten sneller worden.* Hoewel sommige vertaalgeheugens voldoende snel zijn, is dit niet voor alle vertaalgeheugens het geval. Snellere computers met meer intern geheugen kunnen hier de oplossing voor zijn.
- *Vertaalgeheugens zouden meer (of betere) morfologische informatie moeten bevatten.* Hierdoor is het beter mogelijk zinnen te analyseren en een betere schatting te maken van “fuzzy matches”<sup>1</sup>. Bijvoorbeeld samenstellingen zouden door een goed morfologische component herkend kunnen worden.
- *Het invoegen en verwijderen van teksten in het vertaalgeheugen moet makkelijker kunnen.* Nu moet nog vaak met de hand een bestand bijgewerkt worden om stukken tekst en vertalingen aan het vertaalgeheugen toe te voegen of te verwijderen.

Tot nu toe is bij het ontwerpen van vertaalgeheugens veel aandacht besteedt aan zoek- en vergelijkcomponenten. In de huidige vertaalgeheugens werken deze componenten acceptabel, maar vooral de interface naar de gebruiker kan sterk verbeterd worden. Dit zal in de toekomst moeten gebeuren voordat meer vertalers van vertaalgeheugens gebruik zullen maken.

Vertaalgeheugens worden uitgebreid met automatische vertaal-systemen, zodat een redelijk goede automatische vertaling mogelijk is. De databases gebruikt worden door de vertaalgeheugens zijn gedistribueerd op het wereldwijde netwerk aanwezig. Verder wordt veel meer morfologische informatie aan de vertaalgeheugens toegevoegd. Deze zaken lijken nabije toekomst, maar vooral het zoeken in zeer grote vertaalgeheugen-databases en het beheer daarvan is voorlopig praktisch nog niet mogelijk.

<sup>1</sup>Bij een fuzzy match zijn twee segmenten die erg op elkaar lijken, maar niet helemaal hetzelfde zijn. Een voorbeeld van een fuzzy match is twee dezelfde segmenten met verschillende namen erin. De zinnen lijken volledig op elkaar, maar door de verschillende name zijn ze niet exact hetzelfde.

### Automatisch vertalen

Volledig automatisch vertalen met hoge kwaliteit, zonder menselijke interactie tijdens het vertalen, is in slechts enkele projecten geslaagd toegepast. Een voorbeeld hiervan is METEO, waarbij Engelse weerinformatie volledig automatisch vertaald wordt naar het Frans.

Over het algemeen is automatisch vertalen alleen bruikbaar wanneer aan de brontaal zware restricties gesteld wordt of wanneer de in- en uitvoer wordt voor- en nabewerkt. Voor vertalingen van willekeurige teksten is automatisch vertalen niet bruikbaar. De kans is overigens zeer klein dat automatisch vertalen van willekeurige teksten binnen afzienbare tijd algemeen bruikbaar wordt.

Automatisch vertalen kan wel goed gebruikt worden wanneer snel een vertaling nodig is waarbij de kwaliteit niet belangrijk is. Vandaag de dag zijn verschillende automatische vertaalsystemen beschikbaar, waarmee het mogelijk is snel vertalingen van teksten te maken waarbij na vertaling de inhoud van het document duidelijk is, maar zeer zeker niet geschikt is voor publicatie. Verschillende zinsconstructies worden niet goed vertaald en veel woorden zijn niet bekend. Verder hebben automatische vertaalsystemen niet genoeg kennis van de wereld die bij vertalen van belang is. Toch kunnen dit soort systemen nuttig zijn. Automatisch vertalen kan een goede indicatie geven van de inhoud van een tekst, zonder dat de tekst volledig door een vertaler vertaald wordt.

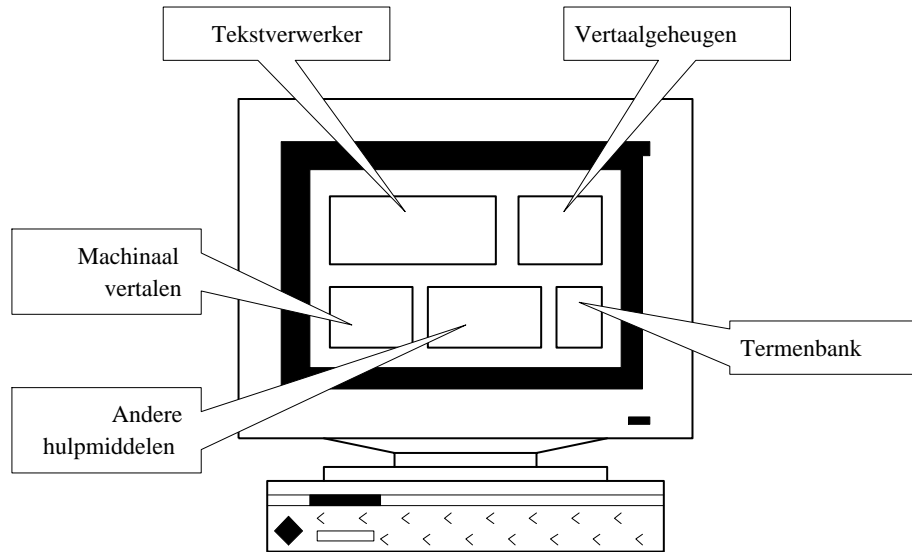
In de toekomst zal automatisch vertalen wel belangrijker worden, maar vooral voor vertalingen van mindere kwaliteit. Hierbij valt te denken aan toepassingen op internet, waarbij informatie met behulp van een online automatisch vertaalsysteem vertaald wordt. Voor vertalingen met hoge kwaliteit zal automatisch vertalen geen goede optie zijn, tenzij de invoer sterk vereenvoudigd is en eventueel de uitvoer bijgewerkt wordt. Voorbeelden hierbij zijn automatische vertalingen van weer- of verkeersinformatie.

Automatische vertaalsystemen zullen kwalitatief veel beter worden. De betere kwaliteit kan gehaald worden door gebruik te maken van een soort zelf regulerend systeem, waardoor na verbeteringen van vertalingen het vertaalsysteem zichzelf bijleert. Door dit soort systemen is een gemakkelijke communicatie mogelijk tussen mensen die verschillende talen spreken. Het vertaalsysteem vertaalt de teksten (en spraak) automatisch naar de goede taal.

### Integratie van verschillende hulpmiddelen

In de toekomst zullen meer en uitgebreidere vertaalomgevingen beschikbaar komen. Deze vertaalomgevingen kunnen verschillende onderdelen hebben, maar de meeste zullen in elk geval een tekstverwerker, een vertaalgeheugen en een termenbank bevatten. Daarnaast kunnen een aantal documentanalyse hulpmiddelen en een automatisch vertaal gedeelte toegevoegd worden. Zie figuur 10.2 voor een overzicht.

Vooral de toevoeging van automatisch vertalen aan een vertaalomgeving kan in de toekomst grote voordelen bieden. De vertaler krijgt dan een mogelijke vertaling aangeboden zelfs bij zinnen die niet bekend zijn in het vertaalgeheugen. In plaats van een volledig automatisch vertaalsysteem wordt automatisch vertalen als hulpmiddel bij het vertalen gebruikt.

**Figuur 10.2** integratie van hulpmiddelen

De integratie van hulpmiddelen wordt zeer ver doorgetrokken tot dat een soort algemeen apparaat ontstaat. Dit apparaat lijkt een beetje op het huidige telefoontoestel, maar heeft een veel uitgebreidere functionaliteit. Het is met het apparaat mogelijk een kwalitatief hoogstaande vertaling met hulpmiddelen of een snelle vertaling met behulp van het vertaalsysteem te maken. Dit apparaat heeft niet alleen de functionaliteit van een telefoon, maar ook van televisie, radio en computer.

### Gebruikersinterface

Vertalers zullen de verschillende vertaalhulpmiddelen (apart of in een complete vertaalomgeving) pas gaan gebruiken als de gebruikersinterface van de programmatuur goed genoeg is. Nu worden langzaam de meeste soorten vertaalhulpmiddelen volwassen, maar de gebruikersinterface is dat duidelijk niet.

In de toekomst zal veel aandacht besteed worden aan de gebruikersinterface van de verschillende vertaalhulpmiddelen. Bij het ontwerpen van een interface zal veel meer uitgegaan moeten worden van de gebruiker. Pas wanneer de gebruikersinterface een stuk beter wordt zullen meer mensen de hulpmiddelen gaan gebruiken.

Bij de gebruikersinterface zal een grotere nadruk komen te liggen op de spraak. De computer kan teksten oplezen, maar ook is het mogelijk teksten in te geven door tegen de computer te praten. Dit alles gaat met dezelfde snelheid als het spreken tegen een persoon. Aan de andere kant wordt virtual reality zeer belangrijk. Virtual reality zorgt ervoor dat het toetsenbord en de muis niet meer nodig zijn.

## 10.2 Publicatie- en vertaaltraject

Het publicatietraject en het vertaaltraject worden nu nog gezien als twee verschillende trajecten. In de nabije toekomst zullen deze trajecten meer en meer in elkaar overvloeien. Voordat dit gebeurt moeten een aantal problemen opgelost worden. Deze problemen zijn vooral van organisatorische aard, maar ook moeten enkele technische problemen opgelost worden.

Naast de bestaande problemen dienen tegelijkertijd enkele veranderingen zich aan. Deze veranderingen kunnen grote invloed hebben op beide trajecten. Om goed met deze veranderingen om te gaan is het nodig de trajecten zo aan te passen dat deze veranderingen gemakkelijk geïntegreerd kunnen worden en zo als voordeel gebruikt kunnen worden.

### 10.2.1 Hulpmiddelen

#### Database publicatie

Bij database publicatie worden delen van documenten in een database opgeslagen. Een nieuw document kan dan samengesteld worden uit de opgeslagen tekstdelen. Hierdoor ontstaan een zeer consistent taalgebruik in de nieuw samengestelde teksten.

Het idee achter het combineren van database publicatie en vertalen is om stukken van documenten in een database op te slaan samen met de vertalingen van dat stuk. Wanneer een nieuw document samengesteld wordt uit de database, is direct de vertaling van het nieuwe document beschikbaar.

Met SGML en XML is het mogelijk inhoudelijke informatie in een document te coderen. Hierdoor is het makkelijker naar inhoud van een tekstdeel te zoeken, maar ook bij het bewerken van en het vertalen van documenten kunnen SGML en XML hulp bieden.

SGML of XML maken het niet direct mogelijk gemakkelijk een stuk tekst terug te vinden in een database. In de toekomst zullen goede technieken ontwikkeld moeten worden om stukken tekst te zoeken of selecteren in een database naar aanleiding van een zoekopdracht. Met dit soort zoekopdrachten moet het mogelijk zijn naar de inhoud van tekstdelen te zoeken. Voordat dit goedwerkende systemen beschikbaar komen moet hier nog veel onderzoek naar verricht worden.

In een andere vorm van database publicatie wordt informatie uit een database gehaald, dat hoeft geen tekstuele informatie te zijn, maar kan bijvoorbeeld technische informatie over bepaalde producten zijn. Deze informatie wordt dan automatisch in een soort standaard document geplaatst. Door dit soort database publicatie is het mogelijk bijvoorbeeld overzichten van beschikbare producten automatisch te produceren.

Veel bedrijven merken dat hun publicatietraject efficiënter kan of zelfs moet worden. Vaak is een betere structurering van documenten een groot gedeelte van de oplossing. Hierdoor zullen in de komende jaren veel bedrijven overstappen op SGML of XML eventueel gecombineerd met database publicatie.

Database publicatie zal in de toekomst erg belangrijk worden, vooral bij documenten die voor een groot gedeelte zijn samengesteld uit delen die hergebruikt kunnen worden in nieuwe documenten. Hoewel tot nu toe database publicatie niet vaak samen met het vertalen van documenten gebruikt wordt, kan hiermee een groot gedeelte van de vertaalkosten en tijd bespaard worden.

Database publicatie zal zeer grote vormen aannemen. Distributed databases zullen op het internationale netwerk aanwezig zijn in veel verschillende talen. Er worden goede zoektechnieken ontwikkeld zodat er snel en gemakkelijk in de verschillende teksten gezocht kan worden.

### Internet

Het internet kan in de toekomst op verschillende manieren een grote rol spelen bij het publiceren. We zullen hier de belangrijkste manieren beschrijven.

Met het internet is een betere communicatie mogelijk. Er kan sneller contact gemaakt worden door middel van bijvoorbeeld e-mail, maar er kan ook interactief gecommuniceerd worden met iemand anders (mogelijk aan de andere kant van de wereld). Het is dus bijvoorbeeld mogelijk sneller documenten te versturen om ze te laten vertalen of gemakkelijk afspraken te maken over oplossingen van problemen.

Betere communicatie zorgt er ook voor dan mensen over de hele wereld samen kunnen werken om bijvoorbeeld nieuwe standaarden of zelfs programmatuur te ontwerpen. Het idee hierbij is dat mensen zelf aanpassingen aan de standaarden en programmatuur kunnen maken en die veranderingen dan opsturen naar de beheerder van de standaard of programmatuur. Een dergelijke werking maakt het mogelijk dat duizenden mensen tegelijkertijd aan een bepaald project werken. Bij een dergelijke constructie kan gemakkelijker een zeer groot project aangepakt worden.

Een ander voordeel van een wereldwijd netwerk als internet is dat informatie voor iedereen beschikbaar worden gesteld. Dat is goed te zien bij het beschikbaar stellen van algemene DTD's<sup>2</sup>. Door het beschikbaar stellen van DTD's hoeft niet iedereen het wiel zelf uit te vinden.

Door het internet wordt het "internationale" gevoel groter. Mensen zullen zich meer een "wereldbewoner" gaan voelen. Daarbij kan de taalbarrière een groot probleem zijn. De vraag naar vertalingen zal dan ook sterk toenemen. Hier zijn twee verschillende oplossingen voor, die waarschijnlijk beide naast elkaar gebruikt zullen worden.

Het document wordt in bijvoorbeeld het Engels beschikbaar gesteld en vertaalsystemen, die on line beschikbaar zullen komen, vertalen de documenten indien nodig. Hiermee zijn snelle vertalingen in veel talen mogelijk, die van mindere kwaliteit zijn. De eerste on line vertaalsystemen komen langzaam maar zeker beschikbaar. Een voorbeeld daarvan is te vinden op

<sup>2</sup>Een DTD, Document Type Description, is een beschrijving van de mogelijke structuur van een document. SGML is een taal waarmee DTD's beschreven kunnen worden.

<http://babelfish.altavista.digital.com/>

Daarnaast kunnen documenten van tevoren in meerdere talen beschikbaar worden gesteld. Deze documenten worden vertaald volgens een vertaaltraject dat erg lijkt op het huidige vertaaltraject. Na dit traject zijn de documenten in verschillende talen beschikbaar, waarbij de vertalingen van hoge kwaliteit zijn.

Aan de andere kant bestaat ook de mogelijkheid dat op het internet de meeste talen zullen verdwijnen en er een soort “wereldtaal” zal ontstaan. Als de huidige trend doorzet, zal de wereldtaal zeer zeker het Engels zijn. Echter, de kans dat dit zal gebeuren is zeer klein. Niet iedereen spreekt Engels en mensen zullen zich afzetten tegen de “ondergang” van hun taal.

In de toekomst zal veel informatie elektronisch beschikbaar worden gesteld. De elektronische informatie zal op verschillende manier toegankelijk worden. Database publicatie is hier van essentieel belang. Het wordt bijvoorbeeld mogelijk te selecteren welke informatie in welk formaat geleverd moet worden. Dat wil zeggen dat de opmaak of het een product waar de informatie over gaat persoonlijk ingesteld kan worden. Ook kan bijvoorbeeld het soort document, dus handleiding, onderhoudshandleiding of verkoopinformatie gekozen worden. Daarnaast kan ook nog geselecteerd worden in welke taal de informatie te zien is. Hoewel dit alles nu al mogelijk is, zal het in de toekomst veel meer gebruikt gaan worden.

<p>Het internet zal uitgebreid worden tot een zeer snel en zeer groot netwerk met veel verschillende mogelijkheden. Het is op verschillende manieren mogelijk contact te zoeken, zoals e-mail en (video-)telefoneren. Kwalitatief goede vertaalsystemen zijn op het netwerk beschikbaar zodat iedereen in zijn eigen taal kan spreken of schrijven. Het is voor de gebruiker niet te zien dat het meerdere computers aan een netwerk zijn, alles is voor de gebruiker afgeschermd. Dit bevordert de gebruiksvriendelijkheid.</p>
--

### 10.2.2 Vraag en aanbod

In de afgelopen paar jaren is de vraag naar vertalingen sterk toegenomen. Er kan een onderscheid gemaakt worden tussen het soort vertalingen dat gevraagd wordt.

Aan de ene kant is er vraag naar vertalingen van mindere kwaliteit, die bijvoorbeeld gebruikt wordt om e-mail of andere persoonlijke correspondentie te vertalen. Dit zal vaak gebeuren met behulp van automatisch vertalen. Het gaat hierbij meer om de inhoud dan om de kwaliteit. Snelheid en gemak zijn bij dit soort vertaling van groot belang, maar ook mag de kostprijs van de vertaling niet te hoog zijn. Tot op heden is de vraag naar dit soort vertalingen nog niet zo groot, maar dat zal in de nabije toekomst veranderen.

Aan de andere kant is er een steeds groter wordende vraag naar vertalingen van hoge kwaliteit. Het soort documenten dat vertaald wordt lopen uiteen van handleidingen tot nieuwsbrieven en brochures. Bij dit soort vertalingen is de kwaliteit van groot belang, dus hiervoor zal voorlopig geen gebruik worden gemaakt van automatische vertalingen, maar wel van andere hulpmiddelen zoals vertaalgeheugens.

De vertalingen van hoge kwaliteit zijn vooral nodig bij bedrijven. Voor deze bedrijven zijn de documenten over het algemeen bijproducten, vaak ondersteunende informatie. Internationaal opererende bedrijven produceren meer en meer ondersteunende informatie in verschillende talen. Hierdoor geeft het bedrijf een betere ondersteuning bijvoorbeeld van de geboden producten. Veel mensen vinden het belangrijk dat ondersteunende informatie in hun eigen taal aanwezig is. Hoewel de publicaties van groot belang zijn, neemt het zelf vertalen en publiceren te veel tijd en moeite in beslag.

De ideale oplossing zou zijn wanneer één bedrijf dit totale traject op zich kan nemen. In dat geval kan de opdrachtgever het document ter publicatie leveren. Het bedrijf dat de publicatie verzorgt, handelt dan het publiceren en het vertalen af. Dat wil niet zeggen dat het bedrijf zelf een vertaalbureau heeft. Het bedrijf kan ook een vertaalbureau inschakelen dat de vertaling verzorgt, maar omdat één bedrijf het totale proces organiseert kan het vertaal- en publicatietraject goed op elkaar afgestemd worden.

Een ander voordeel is dat de opdrachtgever het de totale verwerking van het publicatie- en vertaaltraject uit handen kan geven. De opdrachtgever hoeft niet zelf op zoek naar vertaalbureaus die de documenten vertalen en in een bepaald bestandsformaat terug aanleveren. Ook hoeft de aansluiting naar het publicatietraject niet gemaakt te worden. Er is slechts één bedrijf waar de opdrachtgever mee te maken heeft.

In de toekomst zullen bedrijven die het publiceren en vertalen als een volledig pakket kunnen aanbieden zeer belangrijk worden. Zij bieden de opdrachtgever een compleet publicatie- en vertaaltraject. De opdrachtgever kan het publiceren en vertalen aan één bedrijf uit handen geven en verder kunnen de twee trajecten goed op elkaar aangesloten worden.

<p>Uiteindelijk zal het meeste papier verdwijnen, maar dan moet wel eerst betere hardware ontwikkeld worden. Dit is hardware dat gemakkelijk meegenomen kan worden en een goed leesbare display heeft. Deze nieuwe hardware werkt als een soort boek, waarvan de inhoud veranderd kan worden door bijvoorbeeld een andere cd-rom erin te doen of het apparaat aan het netwerk te koppelen. Uiteraard is er meer mogelijk dan alleen tekst en illustraties. Geluid, video en eventueel geur kunnen ook toegevoegd worden.</p>
--



# Hoofdstuk 11

## Conclusie

Het doel van het integreren van het vertaaltraject in het publicatietraject is dat er een compleet pakket geleverd kan worden waarbij de opdrachtgever een document aanlevert en dat Alfa▼*Base* de rest van het publicatietraject afhandelt. Tot nu toe kwam hier meestal geen vertaalstap aan te pas, maar met de groeiende vraag naar publicaties in meerdere talen is deze uitbreiding een logische stap.

Het publicatietraject kan, zoals het nu is, gemakkelijk uitgebreid worden met een vertaalstap, vooral wanneer een bedrijf wordt ingeschakeld dat vertalingen naar meerdere talen kan afhandelen. Een dergelijk bedrijf organiseert de vertalingen en kan de logistiek van de vertalingen afhandelen, terwijl Alfa▼*Base* de rest van het gehele traject afhandelt.

Het grote, directe probleem bij een de uitbreiding van het huidige publicatietraject met een vertaaltraject is de documentlogistiek. Wanneer een document meer dan één versie krijgt, ontstaat een explosieve groei van documenten en vertalingen.

Dit is uiteraard niet het enige probleem. Voor elke stap in het publicatietraject moet gekeken worden waar de knelpunten zitten met betrekking tot meertalige publicatie. Voorbeelden van problemen zijn de beschikbaarheid van meertalige opmaakprogrammatuur en de beschikbaarheid van meertalige tekstverwerkers.

Een andere manier om meertalige publicaties te produceren is door meer gebruik te gaan maken van database publicaties. Bij een database publicatie wordt veel, zo niet alle, informatie in een publicatie gehaald uit een database. Dit kan duidelijk gestructureerde informatie zijn zoals bijvoorbeeld technische informatie van bepaalde producten, maar het kan ook delen van teksten zijn, die samengevoegd worden. Vooral de laatste soort database publicaties bestaat nog niet erg lang, maar zal in de toekomst waarschijnlijk veel terrein winnen.

Het voordeel van database publicaties is dat vertalingen direct in de database opgeslagen kunnen worden. Deze vertalingen kunnen samen met de andere informatie uit de database gehaald worden, waardoor vertalingen direct geproduceerd worden. Het is nog wel nodig deze vertalingen te laten controleren, dus ook bij dit soort publicaties is de samenwerking met een vertaalbureau noodzakelijk.

Hoe goed Alfa▼*Base* de interne problemen ook heeft opgelost, er blijft altijd het probleem dat meerdere bedrijven tegelijkertijd aan hetzelfde project werken.

Om de verschillende bedrijven goed met elkaar te kunnen laten samenwerken is een goede communicatie tussen de verschillende bedrijven nodig.

## Bijlage A

# Evaluatie van vertaalhulpmiddelen

Het hulpmiddel dat in deze appendix beschreven wordt is apart genomen, omdat het geen vertaalhulpmiddel is, maar een evaluatiehulpmiddel voor vertaalhulpmiddelen. Op dit moment is deze techniek redelijk nieuw en wordt ook nog niet veel gebruikt. Toch is deze techniek hier beschreven omdat het in de toekomst een interessant hulpmiddel kan worden bij het maken van de keuzes in het publicatie- en vertaaltraject.

Bij het kiezen tussen vertaalhulpmiddelen moet gekeken worden welk vertaalhulpmiddel het beste past binnen het traject. Het kiezen van een vertaalhulpmiddel gebeurt op basis van vergelijkingen van de mogelijke hulpmiddelen. Een probleem is nu hoe te bepalen welk vertaalhulpmiddel gekozen moet worden.

Het maken van de keuze is zeer complex. Er zijn namelijk veel verschillende parameters waar rekening mee gehouden moet worden. EAGLES is een onderzoeksgroep die onderzoek heeft gedaan naar evaluatieprocedures. EAGLES is uitgegaan van ISO 9126, dat een evaluatieprocedure beschrijft die gebruikt kan worden bij evaluatie van software in het algemeen.

De evaluatie van vertaalhulpmiddelen op deze manier heeft als voordeel dat er een waardering in de vorm van een getal voor elk gemeten vertaalhulpmiddel beschikbaar komt. Op basis van dat getal kan een keuze gemaakt worden. De waardering die een vertaalhulpmiddel krijgt is afhankelijk van de eisen die aan het vertaalhulpmiddel gesteld worden.

Het verloop van de evaluatieprocedure is als volgt:

**Quality requirements definition** Dit is het definiëren en opschrijven van de behoeften.

**Evaluation preparation** Nu worden de metrieken, de gewenste waarden binnen de metrieken en de definitie van de belangrijkheid van de metrieken onderling gedefinieerd.

**Evaluation procedure** Deze stap valt uiteen in drie tussenstappen:

**Measurement** Het meten met behulp van de metrieken.

**Rating level** De analyse van de gemeten waarden.

**Assessment** verwerking van de uitkomsten van de metingen.

Bij het evalueren wordt uitgegaan van 6 karakteristieken, invalshoeken. Elke karakteristiek heeft zijn eigen metriek. Deze karakteristieken zijn vaak weer onderverdeeld in sub-karakteristieken die weer hun eigen metrieken hebben. Die metrieken worden gecombineerd tot een overkoepelende metriek. De invalshoeken zijn:

1. Functionality,
2. Reliability,
3. Usability,
4. Efficiency,
5. Maintainability,
6. Portability.

Bij het evalueren wordt gebruik gemaakt van een Parametrisable Test Bed. Dit is een systeem dat automatisch de producten kan waarderen. Het PTB maakt daarvoor gebruik van attribuut-waarde paren. Elk te meten product heeft een aantal attribuut-waarde paren, die het product beschrijven. Dit zijn over het algemeen te meten eigenschappen van het te waarderen product, bijvoorbeeld attribuut "aantal gebruikers" heeft waarde "5".

Een PTB is een programma dat

- Gevoed wordt met attributen paren die
  - de te evalueren hulpmiddelen beschrijven,
  - de gebruikers van deze hulpmiddelen beschrijven.
- Een bibliotheek van testmethoden gebruikt.
- De nodige tests uitvoert.
- Een evaluatierapport produceert.

De eisen en mogelijkheden van de hulpmiddelen en de gebruikers van de hulpmiddelen worden beschreven met behulp van attributen. Dan wordt een aantal tests gekozen, waarmee de waarden van de attributen gemeten kan worden. Voor het evaluatierapport worden de attribuut-waarde paren vergeleken met de gewenste waarden van de metrieken. Daaruit is het dan mogelijk een uiteindelijke waarde voor het product te berekenen.

Het grote voordeel van een dergelijke aanpak is dat er zeer gestructureerd te werk kan worden gegaan. Wanneer het traject zorgvuldig gevolgd wordt, zal uiteindelijk een goede evaluatie van de producten gemaakt worden, waarmee een goede keuze gemaakt kan worden.

De moeilijkheid bij een evaluatie als deze is de keuze van de attributen. Ze moeten het te evalueren product goed beschrijven, maar ze moeten ook nog meetbaar zijn. Ook moet aangegeven worden hoe belangrijk bepaalde eigenschappen van het product zijn en welke waarden per attribuut gewenst zijn.

# Bibliografie

- [AJBB] O. Anderson, A. Jakobsen, J. Bøegh, and C.B. Bagge. The microscope standardization context. DELTA Danish Electronics, Light and Acoustics, Hørsholm, Denmark.
- [And89] D. Anderson. *A Guide to Information Sources for the Preparation, Editing and Production of Documents*. Gower Publishing Company Ltd., Hants, UK, 1989.
- [Ben97] M. Benis. Rates charged by interpreters and translators working using the internet/compuserve in 1997. 1997.
- [CJ96] M. Colby and D.S. Jackson. *Using SGML*. Que Corporation, Indianapolis, USA, 1996.
- [Col93] H. Collier. *Strategies in the Electronic Information Industry-A guide for the 1990s*. Infornotics Ltd., UK, 1993.
- [Con96] The Unicode Consortium. *The Unicode Standard, Version 2.0*. Addison-Wesley Developers Press, Reading-Massachusetts, USA, 1996.
- [dEEP95] Eagles document EAG-EWG-PR.2. Eagles-evaluation of natural language processing systems. 1995.
- [Den95] G. Dennet. *Translation Memory: Concept, products, impact and prospects — Major project report*. South Bank University, London, UK, 1995.
- [Gmb97a] Trados GmbH. *Trados Translator's Workbench for Windows-User's Guide*. Trados GmbH, Stuttgart, Germany, 1994–1997.
- [Gmb97b] Trados GmbH. *Trados Translator's Workbench for Windows-Workflow Manual*. Trados GmbH, Stuttgart, Germany, 1997.
- [Gor] I. Gordon. *Letting the CAT out of the bag-or was it MT?* Trados UK Ltd., Kettleshulme, UK.
- [Heya] M. Heyn. *Integrating machine translation into translation memory systems*. Trados.
- [Heyb] M. Heyn. *Present and Future needs in the CAT-World*. Trados.
- [Hut96] J. Hutchings. Computer-based translation systems and tools. *ELRA Newsletter*, 1(4), December 1996.

- [Kaw93] Z. Kawasaki. *Organizational Use of Machine Translation Systems*. In [Nir93], 1993.
- [Law82] V. (ed) Lawson. *Practical Experience of Machine Translation*. North-Holland Publishing Company, Amsterdam, Holland, 1982.
- [New98] EUR-OP News. Europe's tower of babel-colette flesch talks about the role of the translation service. <http://europa.eu.int/comm/opoce/en/opnews/1914-05-1998>, 1998.
- [Nir93] S. Nirenburg. *Progress in machine translation*. IOS Press, Amsterdam, Holland, 1993.
- [NMC] E.H. 3rd Nyberg, T Mitamura, and J.G. Carbonell. *Evaluation Metrics for Knowledge-based Machine Translation*. Carnegie Mellon University, Pittsburgh, USA.
- [Noe95] P.L. Noerr. *Character sets and UNICODE*. Information Management and Engineering Ltd., London, UK, 1995. <http://www.ua.ac.be/KB/pn/pnoerr.txt>.
- [Raya] E.S. Raymond. The cathedral and the bazaar. <http://earthspace.net/esr/writings/cathedral-bazaar/>.
- [Rayb] E.S. Raymond. Homesteading the noosphere. <http://earthspace.net/esr/writings/homesteading/>.
- [RB] C. Rudat and D. Brockmann. *Translation Memory-Systeme: Dokumentenmanagement für Übersetzungen*. Trados GmbH, Stuttgart, Germany.
- [RM80] J. Raben and G. Marks. *Data Bases in the Humanities and Social Sciences*. North-Holland Publishing Company, Amsterdam, Holland, 1980.
- [Rya93] J.P. Ryan. *Machine Translation: Matching Reality to Expectations*. In [Nir93], 1993.
- [Sch86] J.J. Schoorl. *De computer als vertaler*. Boompers drukkerijen BV, Meppel, Holland, 1986.
- [SGa] Germany Star GmbH, Böblingen. *TermStar 2.6*. Star AG, Stein am Rhein, Switzerland.
- [SGb] Germany Star GmbH, Böblingen. *Transit 2.6*. Star AG, Stein am Rhein, Switzerland.
- [sI] samenvatting ISO/IEC 9126. ISO/IEC 9126-information technology-software quality characteristics and metrics.
- [Slo88a] J. Slocum. *A Machine(-Aided) Translation Bibliography*. In [Slo88b], 1988.
- [Slo88b] J. Slocum. *Machine translation systems*. Cambridge University Press, Cambridge, UK, 1988.

- 
- [Slo88c] J. Slocum. *A Survey of Machine Translation: its History, Current Status, and Future Prospects*. In [Slo88b], 1988.
- [Sne79] B.M. (ed) Snell. *Translating and the Computer*. North-Holland Publishing Company, Amsterdam, Holland, 1979.
- [SON93] K. Sakurai, M. Ozeki, and Y. Nishihara. *MT Application for the Translation Agency*. In [Nir93], 1993.
- [Tec95] Cap Volmac Advanced Technology Services Team Support Technology. *Controlled languages automatic document translation*. 1995.
- [TEM97] TEMAA. *TEMAA D16 — Final Report*. TEMAA document D16, 1997.