

Actieplan voor het Nederlands in taal- en spraaktechnologie

Notitie t.b.v. het Nederlands-Vlaams platform voor het Nederlands in taal- en spraaktechnologie

1. Inleiding

- 1.1 *Strategisch belang*
- 1.2 *Knelpunten*
- 1.3 *Kansen*

2. Een platform voor het Nederlands in taal- en spraaktechnologie

- 2.1 *Waarom een platform?*
- 2.2 *Doelstellingen van het platform*
- 2.3 *Instellingstermijn*
- 2.4 *Structuur van het platform*
 - a) TST-platform
 - b) Adviesorgaan
 - c) Werkgroepen ad hoc
 - d) Secretariaat
- 2.5 *Taken van het platform*

3. Actieplan

- Actielijn A: uitbouw van de 'makel en schakel' functie.*
- Actielijn B: plan voor versterking van de digitale taalinfrastructuur*
- Actielijn C: uitwerken standaarden en evaluatiecriteria*
- Actielijn D: plan voor beheer, onderhoud en distributie*

4. Financiering

5. Planning

1 Inleiding

1.1 Strategisch belang

De verregaande integratie tussen informatie- en communicatietechnologie (ICT) hebben geleid tot een maatschappij waarin informatie een centraal element is geworden. Informatie- en communicatietechnologieën (ICT) zijn van cruciaal belang voor het toekomstige concurrentievermogen van alle takken van industrie, handel en dienstverlening. De Europese Commissie verwacht dat de elektronische handel tal van mogelijkheden biedt op het gebied van economische groei, werkgelegenheid en innovatie. Bovendien zou de concurrentiepositie van de EU er sterk door verbeterd worden.

Bij veel ICT-toepassingen vormen taal- en spraaktechnologie (TST) een essentieel onderdeel. Dank zij TST kan de gebruiker zich van natuurlijke taal bedienen om te communiceren met de computer.

- *Taaltechnologie* is niet meer weg te denken als onderdeel van de tekstverwerkeromgeving (grammatica- en spellingcheckers, synoniemenlijsten). Automatische vertaalsystemen dragen op talloze terreinen bij aan het creëren van een meertalige informatiemaatschappij: ze zijn instrumenteel bij het handhaven van de taalkundige verscheidenheid binnen de instellingen van de Europese Unie, zijn onontbeerlijk bij het vertalen van gebruikshandleidingen, ingebouwd in zoekmachines (bijvoorbeeld Altavista) stellen ze de Internetgebruiker in staat te zoeken en te lezen in anderstalige websites. Voor de zichtbaarheid van een taalgebied is het van wezenlijk belang om in deze systemen vertegenwoordigd te zijn.
- *Spraaktechnologie* is essentieel voor de overgang van muisgestuurde naar spraakgestuurde software: automatische informatiediensten, interactieve leermethodes, dicteersystemen, spraakgestuurde apparaten, tekst-naar-spraak omzetteren voor visueel gehandicapten, maken gebruik van spraaksynthese en spraakherkenning.

Naarmate ICT meer deel uitmaakt van het alledaags leven van de gebruiker wordt het belangrijker dat die gebruiker zich van de eigen taal kan bedienen. Alleen op die manier kan men garanderen dat alle burgers volwaardig kunnen participeren in de informatiemaatschappij. Elke taal die een volwaardige positie wil innemen in de meertalige informatiemaatschappij moet gelijke tred houden met de ontwikkelingen in en toepassingen van taal- en spraaktechnologie. Het is daarom noodzakelijk dat de voorzieningen die essentieel zijn voor het ontwikkelen van dergelijke toepassingen aanwezig zijn en kwalitatief voldoen. De Europese Commissie wil bijdragen aan het creëren van een meertalige taalinfrastructuur om zo de meertalige communicatie te garanderen. Elk taalgebied is in Europa echter volgens het subsidiariteitsprincipe verantwoordelijk voor de opbouw van een eigen digitale taalinfrastructuur. De zorg voor taal- en spraaktechnologie staat dan ook op de agenda van verschillende Europese taalgebieden. Ook het Nederlandse taalgebied moet inspanningen verrichten om een volwaardige speler te blijven in een meertalig Europa.

Voor het ontwikkelen van succesvolle toepassingen en producten zijn basisvoorzieningen nodig zoals corpora, elektronische woordenboeken, terminologiebanken, automatische ontleders enz. Het ontwikkelen van deze hulpbronnen gaat de draagkracht van het bedrijfsleven (met name van het MKB / de KMO) te boven. Zelfs in grotere taalgebieden is een extra inspanning van de overheid vereist.

1.2 Knelpunten

Recent werden diverse onderzoeken uitgevoerd om de stand van zaken in Nederland en Vlaanderen op het gebied van taal- en spraaktechnologie in kaart te brengen.

- Binnen het (Europese) EUROMAP-project werden in Nederland en Vlaanderen **markstudies** verricht onder actieve spelers op het gebied van taal- en spraaktechnologie waaronder beleidsmakers, onderzoekers, product- en dienstenaanbieders en gebruikersorganisaties. Doel van de studie was inzicht te verkrijgen in de marktmogelijkheden en in de sterke en zwakke (onderzoeks)componenten.
- In opdracht van de Taalunie werd een onderzoek uitgevoerd naar *de positie van het Nederlands* in Taal- en Spraaktechnologie (rapport Bouma & Schuurman, 1998). Er werd bekeken over welke voor TST benodigde basismaterialen het Nederlands beschikt, welke organisatiestructuren er in Nederland en Vlaanderen op dit gebied bestaan, hoe onderwijs en O&O georganiseerd zijn en welke bedrijven op dit terrein actief zijn.
- Economische Zaken Nederland liet een studie uitvoeren om vast te stellen welke technologiegebieden in de komende tien jaar waarschijnlijk van strategisch belang zullen zijn voor het Nederlandse bedrijfsleven en te onderzoeken of er voldoende kennis wordt opgebouwd op deze gebieden. Uit deze zgn. **TechnologieRadar**

blijkt dat taal- en spraaktechnologie een belangrijke rol speelt bij een aantal van deze voor de toekomst strategisch belangrijke technologieën¹.

Naar aanleiding van deze onderzoeken werd in het kader van de TechnologieRadar door het Ministerie van Economische Zaken in samenwerking met het Ministerie van de Vlaamse Gemeenschap, TNO, NWO, Senter en de Nederlandse Taalunie op 3 februari 1999 de workshop “**Taal- en spraaktechnologie in Nederland en Vlaanderen: naar meer samenwerking?**” georganiseerd. De deelnemers uit het bedrijfsleven en de kennisinfrastructuur bevestigden het bestaan van de knelpunten die in de onderzoeken werden gesignaleerd. De belangrijkste oorzaak voor deze knelpunten is het feit dat taal- en spraaktechnologie zich op het snijvlak van diverse gebieden bevindt waardoor er onvoldoende samenwerking is en het terrein ondoorzichtig is.

Onvoldoende samenwerking

Het is niet duidelijk wie verantwoordelijk is voor de opbouw van een digitale taalinfrastructuur voor het Nederlands. De industrie wil wel gebruik maken van TST-basismaterialen om er toepassingen van te maken maar wil of kan die basismaterialen niet zelf ontwikkelen. De wetenschap wil wel onderzoek doen op het terrein van de taal- en spraaktechnologie maar het eigenlijke ontwikkelen van basisgegevens als corpora behoort niet tot haar taakstelling. Het financieren van het ontwikkelen van taal- en spraaktechnologische materialen behoort ook niet tot de kerntaak van ministeries en departementen. Als het gebeurt is het ook hier op ad-hoc basis (bvb. het corpus gesproken Nederlands).

Deze versnippering bemoeilijkt het voeren van een coherent beleid en maakt het veld ondoorzichtig voor geïnteresseerde partijen.

Bedrijfsleven en kennisinfrastructuur vaak onbekenden voor elkaar

Bedrijfsleven en kennisinfrastructuur weten elkaar moeilijk te vinden; interesses en invalshoeken sporen niet. Vraag (bedrijfsleven) en aanbod (kennisinfrastructuur) zijn niet goed op elkaar afgestemd: het bedrijfsleven heeft onvoldoende zicht op wat de kennisinfrastructuur te bieden heeft, de kennisinfrastructuur ontwikkelt ‘producten’ waar het bedrijfsleven niet op zit te wachten.

Tekort aan goed opgeleide taal- en spraaktechnologen

Door inbedding van de opleiding in ofwel de taalwetenschap (letterenfaculteit) ofwel de technische opleiding (informaticafaculteit) wordt het uitbouwen van een multidisciplinair curriculum bemoeilijkt en is de instroom van studenten met interesse voor de beide aspecten (taal en technologie) te beperkt om aan de vraag van het bedrijfsleven te voldoen.

De samenwerking tussen taaltechnologen (veelal komend uit de taalhoek) en spraaktechnologen (veelal komend uit de ingenieurshoek) is slechts recent tot stand gekomen.

Het Nederlands heeft een lage plaats op de prioriteitenladder van het bedrijfsleven

Investerings voor Nederlandstalige toepassingen zijn dezelfde als voor toepassingen in grotere talen, de afzetmarkt is echter kleiner. Voorts zijn er maar weinig leveranciers van op het Nederlands gerichte technologie.

1.3 Kansen

Naast deze knelpunten kunnen ook sterktes geconstateerd worden. Een aantal Nederlandse en Vlaamse producten en bedrijven hebben zich een vooraanstaande positie op de wereldmarkt verworven. Zij kunnen bijdragen aan de totstandkoming van de benodigde Nederlandstalige producten en –diensten. Voorts hebben zij belangrijke uitstralingseffecten naar het overig geïnteresseerde bedrijfsleven en de kennisinfrastructuur.

Daarnaast hebben de Nederlandse en Vlaamse kennisinfrastructuur wereldwijd een goede naam op een aantal terreinen van het onderzoek. Recent gestarte grootschalige Nederlands-Vlaamse projecten hebben bijgedragen tot concrete, hechte samenwerking tussen een groot aantal wetenschappers uit verschillende richtingen uit beide landen.

Het gaat er nu om de onderlinge samenwerking te verbeteren en de reeds aanwezige kennis en kunde aan te wenden voor de totstandkoming van de benodigde infrastructuur voor het Nederlands. Alleen op die manier kunnen de knelpunten worden weggenomen en kan de Nederlandse en Vlaamse ambitie voor een volwaardige participatie van het Nederlands in de meertalige informatiemaatschappij worden waargemaakt.

¹ m.n. de gebieden Data- en kennissystemen en Multimedia en interactie.

2 Een platform voor het Nederlands in taal- en spraaktechnologie

2.1. Waarom een platform?

Om het Nederlands een volwaardige rol te laten spelen in de meertalige informatiemaatschappij is het belangrijk dat een inspanning wordt verricht om de ontbrekende basismaterialen te ontwikkelen. Dit is een kostbare onderneming die de draagkracht van individuele betrokkenen te boven gaat. Samenwerking tussen de verschillende actoren (beleid, kennisinfrastructuur en bedrijfsleven) in Nederland en Vlaanderen is geboden. Samenwerking kan leiden tot meerwaarde/voordeel op allerlei terreinen:

- door gezamenlijke prioriteitstelling kunnen financiën zinvol worden aangewend;
- informatie-uitwisseling voorkomt dubbel werk en biedt meer kwaliteitsgaranties door onderlinge afspraken over standaarden en multifunctionaliteit;
- synergie door het samen ontwikkelen van bepaalde overlappende elementen en extra impuls om multifunctionele materialen te ontwikkelen (niet enkel voor een specifiek doel) en te onderhouden;
- het biedt mogelijkheden tot informatie-uitwisseling over geschikte experts voor uitvoering en evaluatie.

Een platform

- biedt een structureel forum voor dialoog en overleg tussen de verschillende actoren (beleid, kennisinfrastructuur en industrie) in Nederland en Vlaanderen;
- maakt het mogelijk gezamenlijk prioriteiten te stellen, voorstellen voor beleid te formuleren en acties te ontwikkelen met betrekking tot versterking van de positie van het Nederlands in taal- en spraaktechnologie;
- kan als aanspreekpunt, wegwijzer fungeren doordat er een centraal overzicht is van wie wat waar ontwikkelt of ontwikkeld heeft;
- verhoogt de slagkracht nodig om de belangen van het Nederlandstalig gebied in een meertalig Europa te waarborgen.

Het Comité van Ministers van de Nederlandse Taalunie onderschrijft het belang van een adequate digitale taalinfrastructuur voor het Nederlands en besloot op basis van de aanbevelingen in het rapport *De positie van het Nederlands in taal- en spraaktechnologie* tot de oprichting van een Nederlands-Vlaams platform voor het Nederlands in taal- en spraaktechnologie. Het platform wordt officieel opgericht op 19 april 1999.

2.2 Doelstellingen van het platform

Het platform stelt zichzelf ten doel de nodige inspanningen te verrichten om

1. de positie van het Nederlands in taal- en spraaktechnologie te bevorderen zodat het Nederlands een volwaardige rol kan spelen in de informatiemaatschappij en het Nederlands een volwaardige speler blijft in een meertalig Europa;
2. de juiste condities en een adequaat kader te scheppen voor succesvol beheer en onderhoud van met overheidsgelden ontwikkelde basismaterialen;
3. instrumenteel te zijn bij verdergaande samenwerking tussen kennisinfrastructuur en bedrijfsleven op het gebied van taal- en spraaktechnologie;
4. instrumenteel te zijn bij het tot stand brengen van samenwerking op Europees niveau op relevante terreinen binnen taal- en spraaktechnologie;
5. een netwerk op te bouwen, zowel elektronisch als in de persoonlijke sfeer dat vraag naar en aanbod van kennis, producten en diensten effectief bij elkaar brengt.

2.3 Instellingstermijn

Het platform wordt in eerste instantie ingesteld voor een periode van vijf jaar (1999-2004). Na drie jaar zal het platform de eigen werking en resultaten evalueren zodat het werkplan tussentijds kan worden bijgesteld. Indien na afloop van deze periode blijkt dat het platform in een duidelijke en continue behoefte voorziet zal worden overwogen de instellingstermijn te verlengen. Op basis van een evaluatierapport zullen taken, doelstellingen en structuur dan opnieuw worden gedefinieerd.

2.4 Structuur van het platform

a) Nederlands-Vlaams platform voor het Nederlands in taal- en spraaktechnologie (TST-platform)

In het TST-platform zijn de beleidsorganisaties vertegenwoordigd die in Nederland en Vlaanderen verantwoordelijk zijn voor het beleid t.a.v. taal- en spraaktechnologie. Het is niet de bedoeling een nieuwe structuur te creëren maar bestaande structuren te coördineren. Het platform moet daarom gezien worden als een flexibele constructie waarbinnen de partners de agenda's met betrekking tot taal- en spraaktechnologie op elkaar afstemmen en al dan niet besluiten nieuwe onderwerpen op een gezamenlijke agenda te plaatsen. Naast de Nederlandse Taalunie, werken de volgende organisaties samen in het TST-platform: in *Nederland*: het Ministerie van Onderwijs, Cultuur en Wetenschappen, het Ministerie van Economische Zaken, NWO en Senter; in *Vlaanderen*: het Ministerie van de Vlaamse Gemeenschap/AWI, het IWT en FWO-Vlaanderen. De noodzakelijke samenwerking met het veld zal op een andere wijze worden vormgegeven: door het creëren van een permanent adviesorgaan en door het instellen van werkgroepen ad hoc.

b) Nederlands-Vlaams adviesorgaan

Dit Nederlands-Vlaams adviesorgaan bestaat uit een netwerk van TST-experts uit bedrijfsleven en kennisinfrastructuur. Het werkterrein van deze adviescommissie is de materiële infrastructuur. Uit dit netwerk kan geput worden naargelang van specifieke vragen. Het adviesorgaan wordt geraadpleegd bij de evaluatie van bestaande producten en hulpmiddelen en bij de evaluatie van nieuwe aanvragen voor de adaptatie van de bestaande infrastructuur of de opbouw van nieuwe basisvoorzieningen en toepassingen. Binnen het Corpus gesproken Nederlands is reeds een uitgebreid netwerk opgebouwd. Dit zal verder worden uitgebouwd.

c) Nederlands-Vlaamse werkgroepen ad hoc

Voor het uitwerken van specifieke opdrachten zullen werkgroepen ad hoc worden ingesteld. De werkgroepen bestaan uit TST-experts uit kennisinfrastructuur en bedrijfsleven (en eventueel vertegenwoordigers van het beleid). Deze werkgroepen worden in het leven geroepen voor een welbepaalde periode en krijgen een welbepaalde taak (bvb. opstellen van een werkplan, opstellen van een blauwdruk, uitwerken van standaarden en evaluatiecriteria).

d) Secretariaat

Voor de goede werking van het geheel is structurele ondersteuning nodig in de vorm van een secretariaat. Tot het takenpakket van het secretariaat behoren naast louter administratieve taken ook

- het actueel houden van de inventarissen (bvb. middels een gestandaardiseerde enquête of middels een elektronisch meldingspunt) zodat alle geïnteresseerden via Internet actuele gegevens krijgen over wie waarmee bezig is en welke materialen waar beschikbaar zijn;
- het fungeren als informatieve helpdesk voor technische en informatieve vragen; voor inhoudelijke suggesties en vragen kan worden doorverwezen naar andere, bestaande helpdesks van kennisinstututen en bedrijven;
- het organiseren van werkbijeenkomsten en informatiedagen in gezamenlijk overleg tussen beleid, kennisinfrastructuur en bedrijfsleven.

Op die manier kan het platform-secretariaat fungeren als nationale en internationale draaischijf tussen beleid, bedrijfsleven en kennisinfrastructuur met betrekking tot het Nederlands in taal- en spraaktechnologie

Voor concrete projecten treedt het secretariaat op als coördinator en wordt in overleg met de projectpartners extra capaciteit elders gezocht.

2.5 Taken van het platform

Het platform wil instrumenteel zijn voor samenwerking op materieel en immaterieel niveau:

a) op materieel niveau

- verbeteren en versterken van de materiële infrastructuur (de zgn. basismaterialen nodig voor een adequate digitale taalinfrastructuur voor het Nederlands):
 - opstellen van een plan met prioriteitenlijst en een kostenraming (zie Actielijn B);
 - opstellen van een plan voor het beheer, onderhoud en ter beschikking stellen van TST-materialen (zie Actielijn D);
 - laten uitwerken van een evaluatie-instrumentarium (zie Actielijn C);
 - entameren concrete acties.

b) immaterieel niveau (zie Actielijn A)

- fungeren als aanspreekpunt op het gebied van taal- en spraaktechnologie voor de overheid en andere partijen (informatie verzamelen en verstrekken)
 - monitorfunctie in de vorm van een continue informatievergarende en informatieverstreckende helpdesk;
- ‘makel en schakel’ functie
 - draaischijf van contacten tussen overheid, bedrijfsleven en kennisinfrastructuur in Nederland en Vlaanderen;
 - bevorderen van Nederlands-Vlaamse samenwerking (expertisecentra, industriële kenniscentra, dataleveranciers enz.);
 - bevorderen van participatie aan Europese projecten en initiatieven (Vijfde kaderprogramma).
- initiëren voorlichtingsactiviteiten
 - organiseren van werkbijeenkomsten, colloquia, informatiedagen;
 - stimuleren van *awareness*-acties.

3 Actieplan

Het doel van het platform is de verder uitbouw van een adequate digitale taalinfrastructuur voor het Nederlands zodat het bedrijfsleven de nodige toepassingen ontwikkelt die moeten garanderen dat de burgers in Nederland en Vlaanderen zich van de eigen taal kunnen bedienen in hun communicatie binnen de informatiemaatschappij en het Nederlandse taalgebied een volwaardige speler blijft in een meertalig Europa. De partners binnen het platform zullen hun inspanningen op dit gebied bundelen.

Een aantal gesignaleerde knelpunten worden voornamelijk buiten het Actieplan gehouden. Zo is de academische inbedding en de invulling van het curriculum van de opleiding taal- en spraaktechnologie onderwerp van nationaal beleid, de afstemming hierover tussen Nederland en Vlaanderen zou eventueel wel op agenda van het platform kunnen worden geplaatst.

Om de bestaande en toekomstige resultaten te maximaliseren moeten de juiste randvoorwaarden worden geschapen. Het platform zal hiertoe een aantal kortlopende acties initiëren:

Actielijn A: uitbouwen ‘makel-en-schakel’ functie.

Actielijn B: plan voor de versterking van de digitale taalinfrastructuur

Actielijn C: opstellen standaarden en evaluatiecriteria

Actielijn D: plan voor beheer, onderhoud en distributie

Actielijn A is een ‘horizontale’ actielijn, de resultaten van deze actielijn hebben rechtstreeks effect op de actielijnen B, C en D. Anderzijds versterken de resultaten van de Actielijnen B, C en D het effect van actielijn A. De acties B, C en D moeten parallel verlopen zodat gegevensuitwisseling mogelijk is. De resultaten van deze acties moeten instrumenten bieden voor efficiënt gezamenlijk beleid zoals basisgegevens die als grondslag kunnen dienen bij het beoordelen projectvoorstellen, het formuleren van randvoorwaarden voor projecten en het evalueren van projectresultaten.

ACTIELIJN A: Uitbouw van de makel- en schakelfunctie

doel:

Het gebrek aan samenwerking tussen de diverse partijen wordt als belangrijk knelpunt ervaren. Het platform zal binnen Actielijn A een aantal activiteiten ondernemen om de samenwerking te bevorderen. Het betreft daarbij niet enkel de partijen die zich op dit ogenblik reeds bewegen op het terrein van de taal- en spraaktechnologie maar ook potentieel nieuwe partijen. Voorts wil het platform een rol spelen in de samenwerking tussen Europa en het Nederlandse taalgebied op dit terrein.

beoogd resultaat:

betere samenwerking tussen de diverse partijen,
betere zichtbaarheid van het veld.

werkwijze:

Het platform zal waar mogelijk het belang van de positie van het Nederlands in taal- en spraaktechnologie op de agenda plaatsen. Voorts zal het platform werkbijeenkomsten, colloquia, informatiedagen organiseren. Een klein maar efficiënt secretariaat zal de activiteiten van het platform ondersteunen. Het secretariaat zal ook optreden als een permanente informatievergarende en informatieverstreckende helpdesk en fungeren zo als aanspreekpunt op het gebied van TST voor de overheid en andere partijen (informatie verzamelen en verstrekken).

Actielijn B:**Opstellen van een concreet plan voor de versterking van de digitale taalinfrastructuur****doel:**

Zowel voor het ontwikkelen van succesvolle (commerciële) toepassingen als het verrichten van (strategisch onderzoek) zijn basismaterialen nodig. In 2003 moet reeds een belangrijke vooruitgang zijn geboekt in het ontwikkelen van deze basismaterialen. Het doel van Actielijn B is te definiëren wat de zgn. BLARK (*Basis Language Resources Kit*) voor het Nederlands moet inhouden, welke materialen bij voorrang moeten worden ontwikkeld en welke kosten daaraan verbonden zijn.

beoogd resultaat:

In 2000 moet een prioriteitenlijst (met kostenraming) zijn opgesteld die als richtlijn kan dienen voor het beleid.

internationale samenwerking:

Bij het definiëren van de BLARK zal worden samengewerkt met ELSNET.

werkwijze:

In het rapport *De positie van het Nederlands in taal- en spraaktechnologie* (Bouma & Schuurman 1998) wordt de stand van zaken voor het Nederlands reeds in kaart gebracht en een eerste aanzet gegeven voor een prioriteitenlijst. In het rapport wordt een overzicht gegeven van zowel de *benodigde* als de reeds *bestaande* componenten voor een *minimale* én een *optimale* infrastructuur. In deze lijst zijn nog enkele hiaten (m.n. op het gebied van spraaktechnologie) die opgevuld moeten worden.

Nog in 1999 wordt een werkgroep ad hoc ingesteld die op basis van deze gegevens en verder onderzoek een plan moet opstellen voor de versterking van de materiële infrastructuur van het Nederlands. De volgende stappen moeten daarbij worden gezet:

- beschrijving van reeds bestaande basismaterialen: op basis van een kwalitatieve evaluatie (samenwerking met Actielijn C) moet blijken of wat bestaat ook kwalitatief voldoet en (auteursrechtelijk) beschikbaar is. Dit geeft een duidelijker inzicht in wat tot de reeds bestaande materiële infrastructuur kan worden gerekend. Op basis van de resultaten kan blijken welke materialen geschikt, niet geschikt (bijvoorbeeld niet multifunctioneel of niet beschikbaar) of slechts na adaptatie geschikt zijn als onderdeel van de infrastructuur. Zo ontstaat een realistisch overzicht van de huidige stand van zaken en krijgt men tevens een idee van de kennis- en tijdsinvestering voor het modulair aanpassen en vervolgens beschikbaar maken van bestaande materialen.
- definiëren van een minimale infrastructuur (BLARK);
- opstellen van een lijst van op middellange termijn te ontwikkelen basismaterialen door de reël bestaande basismaterialen af te zetten tegen de minimale infrastructuur;
- opstellen van een prioriteitenlijst;
- opstellen van een kostenraming bij deze prioriteitenlijst.

Voor het opstellen van een prioriteitenlijst, i.e. een lijst van materialen die bij voorrang moeten worden ontwikkeld, is een dialoog nodig tussen bedrijfsleven (wat is nodig?), kennisinfrastructuur (wat is nodig?, wat is mogelijk?) en beleid (hoe faciliteren?). Vanzelfsprekend moet bij de prioritering eerst aan de eisen van de minimale infrastructuur worden voldaan. Toch moet alles wat nu reeds voorhanden is en wat niet tot de minimale infrastructuur behoort, van bij de start bij het geheel worden betrokken.

Actielijn C:**Uitwerken van standaarden en evaluatiecriteria****doel:**

Het opstellen van een lijst met criteria waaraan basismaterialen moeten voldoen.
Het opstellen van standaarden voor de validatie van Nederlandstalige materialen.

beoogd resultaat:

Een modulair evaluatie-instrumentarium waarvan gebruik kan worden gemaakt bij het beoordelen van projectaanvragen en het evalueren van projectresultaten.

internationale samenwerking:

Bij het opstellen van standaarden en evaluatiecriteria zal worden samengewerkt met ELRA en EAGLES. Ook zal gebruik worden gemaakt van de ARPA/DARPA-resultaten.

werkwijze:

Het is niet voldoende dat geconstateerd wordt dat een bepaalde basisvoorziening aanwezig is: de basisvoorziening moet ook voldoen aan eisen van formele en inhoudelijke kwaliteit, beschikbaarheid (rechtenvrij of onder bepaalde voorwaarden), multifunctionaliteit en hergebruik (niet enkel bruikbaar voor de toepassing waarvoor oorspronkelijk ontworpen) en moet gedocumenteerd zijn. In de rapporten van de Taalunie en van Euromap wordt geen aandacht geschonken aan het evalueren van het bestaande materiaal met betrekking tot bruikbaarheid als TST-basisvoorziening. Voor de evaluatie zou men in principe gebruik kunnen maken van de resultaten van verschillende Europese projecten en organisaties (EAGLES, TSNLP, ELSE etc.). Deze resultaten zijn echter niet gebaseerd op medecontrole of op mede-evaluatie van Nederlandstalige producten en hulpmiddelen (zoals ondersteunende software).

Om tot een werkbaar resultaat te komen, moeten de volgende stappen worden doorlopen:

1. instellen van een overkoepelende werkgroep ad hoc met de opdracht per relevant TST-domein een subwerkgroep in te stellen die als taak heeft een maximale set van objectieve criteria te definiëren waaraan elk specifiek onderdeel van een type van basisvoorziening (producten en hulpmiddelen) moet voldoen;
2. inventariseren van bestaande evaluatiehulpmiddelen voor BLARK-onderdelen waarvoor het Nederlands al dan niet als input fungeerde (zoals binnen het EAGLES-project);
3. opstellen van een overkoepelende taxonomie voorzien van een exemplarische invulling;
4. opstellen van algemene controlelijst op metaniveau (taak overkoepelende werkgroep ad hoc) en van verschillende typeafhankelijke controlelijsten bestaande uit maximale sets van eisen en criteria (taak subwerkgroepen);
5. opstellen van minimale controlelijsten die typeafhankelijk zijn (samenwerking tussen de verschillende subgroepen);
6. uitvoeren van de evaluatie van alle beschikbare en in ontwikkeling zijnde TST-basisvoorzieningen (producten en hulpmiddelen);
7. opstellen van een overzicht waarin per geëvalueerde basisvoorziening de resultaten worden weergegeven samen met de gewenste, haalbare opwaardering, gekoppeld aan een geschatte termijn en kostenraming.

De set van criteria kan leiden tot het opstellen van een gelaagde of modulaire controlelijst² die het mogelijk maakt om op een objectieve manier de bestaande materialen te beoordelen, vergelijken en rangschikken. Die

² Enkele vragen die van de controlelijst deel uitmaken, zijn:

1. is een valabele linguïstische beschrijving voorhanden? (b.v. valentiewoordenboek, tagset voor sem. pragm. labeling, vorm: folio, elektronisch, niet, beide)
2. zijn er standaarden beschikbaar?
3. is er gebruik gemaakt van een standaard of van een bepaald consistent systeem?
4. wat zijn de gebruiksmogelijkheden, technisch gezien, in termen van flexibiliteit?
5. hoe wordt de interface ervaren door de verschillende groepen van gebruikers?
6. indien het materiaal niet bruikbaar is, is aanpassing dan kosteneffectief? (expertise, tijd, locatie, integreerbaarheid) en wat zijn de verschillende fasen in het updating- of upgradingsproces?
7. indien het product niet voldoet, is er dan een tool voorhanden die met een nieuwe set gegevens betere resultaten kan creëren? Indien niet, is de bestaande tool adapteerbaar op een kosteneffectieve manier?
8. indien het adapteren van het materiaal en/of de tool niet kosteneffectief is, wat is dan de beste strategie om opnieuw vanaf het begin te beginnen, rekening houdend met de prioriteitenlijst, de beschikbare middelen en de incorporatie in een groter geheel (b.v. Ned. component in bestaand EU-project)
9. indien re-creatie wordt overwogen, wat is dan het procédé volgens welk een kwaliteitsetiket wordt toegekend door het platform?
10. kan de controlelijst worden ingezet voor het toekennen van een inhoudelijk kwaliteitsetiket gegeven door de permanente inhoudelijke adviescommissie? (≠ kwaliteitsetiket van het platform)?
11. welke verwervingsscenario's zijn er?
12. hoe kan het verworven materiaal raadpleegbaar worden gemaakt?
13. hoe (kanalen, licenties) kan de infrastructuur best worden verspreid?

controlelijst kan ook worden ingezet bij het toekennen van een kwaliteitsetiket aan herwerkte en nieuwe producten en hulpmiddelen.

ACTIELIJN D:

Opstellen van een plan voor beheer, onderhoud, terbeschikkingstelling en distributie

doel:

Verschillende organisaties worden in de praktijk geconfronteerd met beheer, onderhoud en distributie van TST-materialen. Instanties financieren vaak enkel het ontwikkelen van materialen en voelen zich niet meer verantwoordelijk na afloop van het project. Materialen die niet worden onderhouden, verminderen echter snel in waarde. Onduidelijke auteursrechtelijke afspraken kunnen exploitatie bemoeilijken. Het doel van Actielijn D is procedures te ontwikkelen die voor diverse projecten en organisaties bruikbaar zijn.

beoogd resultaat:

Blauwdruk voor beheer, onderhoud, terbeschikkingstelling en distributie van materialen die kunnen worden ingezet bij onderwijs en onderzoek en bij de ontwikkeling van producten op het gebied van de taal- en spraaktechnologie.

internationale samenwerking met: ELRA

werkwijze

In 1999 wordt een werkgroep ad hoc ingesteld met als opdracht het opstellen van een blauwdruk voor beheer, onderhoud en distributie van met overheidsgeld ontwikkelde basismaterialen. Het beheer omvat o.a. het treffen van overeenkomsten met de eigenaars van het materiaal over eigendomsrechten en distributie, het eventueel aanpassen van het materiaal zodat het op verschillende computersystemen gebruikt kan worden, het beschikbaar maken van documentatie, het fungeren als helpdesk, het onderhouden en actualiseren van het materiaal.

Combinatie van materialen kan meerwaarde opleveren. In de blauwdruk moet daarom ook aandacht gegeven worden aan manieren waarop organisaties/personen die materialen beheren structureel kunnen samenwerken. Men kan daarbij denken aan een virtueel instituut waarbij de verschillende organisaties/personen een aantal taken gezamenlijk uitvoeren. Het voordeel hiervan is niet enkel verhoogde efficiëntie maar ook een grotere zichtbaarheid van het Nederlandse taalgebied.

De blauwdruk kan aanknopingspunten bieden voor diverse organisaties die met dezelfde problematiek geconfronteerd worden. Ieder hoeft niet voor zich 'het wiel uit te vinden' en bijvoorbeeld (duur) juridisch advies in te winnen. Bij het opzetten van nieuwe projecten kan rekening gehouden worden met de aanbevelingen waardoor in een vroeg stadium rekening wordt gehouden met zaken als auteursrecht, financiering van onderhoud, mogelijkheden voor distributie.

4 Financiering

Voor de uitvoering van concrete projecten om de taaltechnologische infrastructuur van het Nederlands te versterken zouden ad hoc budgetten van verschillende betrokkenen samengevoegd kunnen worden zoals nu gebeurt voor de financiering van het corpus gesproken Nederlands (waarbij de partners zeggenschap houden over de eigen bijdrage). In eerste instantie worden de diverse in aanmerking komende beleids- en financieringsinstrumentaria van de verschillende partners geïnventariseerd. Op deze wijze kan duidelijker in beeld worden gebracht welke activiteiten onder welke voorwaarden voor welke programma's in aanmerking komen.

De partners worden geacht financieel bij te dragen aan de werking van het platform en het uitvoeren van de actielijnen zoals gedefinieerd in het Actieplan.

5 Planning

- eerste rondetafelgesprek met beoogde partners in het platform (Den Haag 8 februari 1999);
- opstellen van beleidsnota met doelstellingen, taken en Actieplan van het platform op basis van de besluiten van het 1^e rondetafelgesprek (februari-maart 1999);
- tweede voorbereidend gesprek met partners van het TST-platform (Antwerpen, 10 maart 1999);
- officiële instelling platform tijdens *Flanders Technology International* (Technoland Gent, april 1999);
- begin werkzaamheden platform (vanaf mei 1999)
 - instellen van Vlaams-Nederlandse werkgroepen ad hoc, bvb.:
 - a) werkgroep opstellen prioriteitenlijst materiële infrastructuur en financieel plan (1999-2000),
 - b) werkgroep vaststellen evaluatiecriteria (1999-2001),
 - c) werkgroep blauwdruk beheer, onderhoud en distributie van bestaande materiële infrastructuur (1999-2000);
 - start opbouw adviesnetwerk met TST-experts uit bedrijfsleven en kennisinfrastructuur;
 - vervullen van monitorfunctie in de vorm van een informatievergarende en -verstrekende elektronische helpdesk;
 - acties ter bevordering van de Nederlands-Vlaamse en Europese samenwerking;
- mei 2000: presentatie van de activiteiten van het platform aan een groter publiek tijdens de Talenbeurs in Gent, dit is een grootse manifestatie in het kader van de Keizer Karelviering;
- 2000: resultaten actielijn B (prioriteitenlijst) en D (blauwdruk voor het beheer, onderhoud en distributie);
- 2001: resultaten actielijn C: standaarden en evaluatiecriteria);
- 2002: tussentijdse evaluatie;
- 2004: afronding en evaluatie activiteiten eerste termijn.

Lijst van afkortingen

BLARK: Basis LAnguage Resources Kit
DARPA: Defense Advanced Research Projects Agency (VS)
EAGLES: Expert Advisory Group on Language Engineering Standards
ELRA: European Language Resources Association
ELSE: Evaluation in Language and Speech Engineering
ELSNET: European Network in Language and Speech
TSNLP: Test Suites for Natural Language Processing

Op dit ogenblik lopen reeds de volgende, gezamenlijke projecten:

Corpus gesproken Nederlands

Nederlands-Vlaams project met als doel het ontwikkelen van een corpus van 10 miljoen woorden gesproken Nederlands. Het programma werd eind 1997 officieel goedgekeurd door de Nederlandse en Vlaamse regering. De totale kosten worden begroot op 10.000.000 gulden (183.055.000 BEF volgens de wisselkoers op 8/4/1999). De Vlaamse regering stelde een bedrag van 63 miljoen BEF beschikbaar voor het Vlaamse deel van het project. In Nederland stelde het Ministerie van Onderwijs, Cultuur en Wetenschappen (OCenW) 3 miljoen gulden beschikbaar. NWO kende in totaal 3 miljoen gulden toe (2 miljoen vanuit het stimuleringsfonds en 1 miljoen vanuit het Internationaliseringsfonds. Het beheer van het Vlaamse budget is in handen van AWI, het beheer van het Nederlandse budget is in handen van NWO.

NL-Translex

Europees MLIS-project met als doel de ontwikkeling van Nederlandstalige componenten in systemen voor automatisch vertalen van teksten. Het project is gericht op de ontwikkeling van componenten in systemen voor machinevertalen die vertalingen mogelijk maken met het Nederlands als bron- én doeltaal. De andere beoogde talen in de vertaalparen zijn in volgorde van prioriteit het Engels, Frans en Duits. Het project werd einde 1998 goedgekeurd door de Europese Commissie
Partners: naast de Nederlandse Taalunie (als coördinator) in Nederland: Ministerie van Onderwijs, Cultuur en Wetenschappen, Ministerie van Economische Zaken, Ministerie van Buitenlandse Zaken (Hoofdafdeling Vertalingen), vertaalbureau van de Sociale verzekeringsbank; in Vlaanderen: IWT en de Kanselarij van het Ministerie van de Vlaamse Gemeenschap. Ook de Vertaaldienst van de Europese Commissie is bij het project betrokken. Het totale budget over de hele projectperiode 1.250.000 euro. Die worden als volgt verdeeld: de Partners Nederlandse Taalunie: 400.000 euro, MLIS: 450.000 euro; van de systeemaanbieder wordt een bijdrage van minimaal 400.000 euro gevraagd.