

# De ketenbenadering<sup>1</sup>

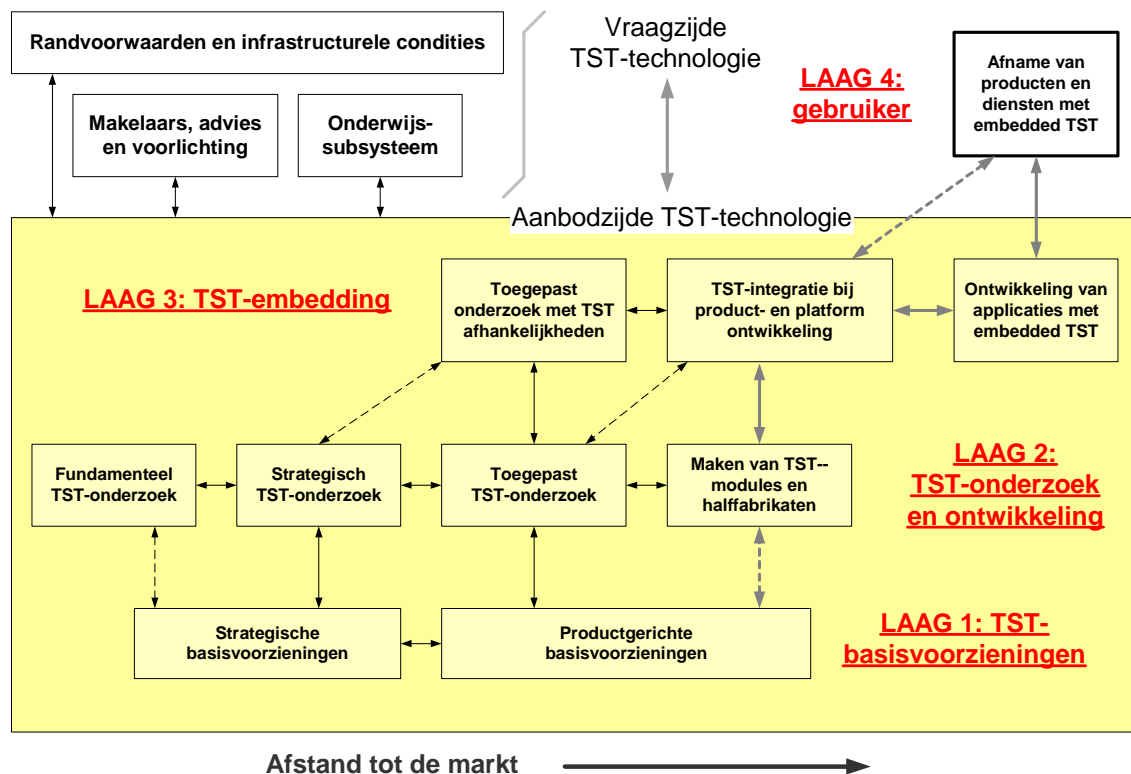
In de *Technologieverkenning Nederlandstalige taal- en spraaktechnologie* van M&I/Partners wordt uitgebreid uiteengezet hoe het TST-innovatieproces kan worden beschouwd als een gelaagde keten met vier lagen (zie figuur). Op al deze lagen zijn verschillende soorten bedrijven actief. Voor zover het gaat om de basistaalvoorzieningen (laag 1) betreft het private partijen die bestanden bezitten gericht op een specifiek marktsegment, maar ook woordenboekfabrikanten. In het TST-ontwikkelingsproces (laag 2) zijn twee typen bedrijven te onderscheiden: de TST-componentenbouwers die halffabrikaten vervaardigen die ze weer doorleveren aan platformbouwers en integratoren. Daarnaast zijn ook applicatieontwikkelaars actief die platformen en TST-componenten aankopen en daarmee in eigen beheer of in opdracht producten en/of diensten maken. Tot slot, in het commercialisatieproces (lagen 3 en 4) zijn de volgende actoren actief: de platformbouwers, die hun eigen commerciële producten/diensten willen verbeteren door het inbouwen van TST, de distributeurs van producten en de uitbaters van diensten (o.a. banken, de overheid, call centres, etc.).

Soms komt dit onderscheid echter vrij kunstmatig over en is er in de realiteit sprake van een diffusie tussen de lagen, aangezien sommige bedrijven enerzijds taal- en spraaktechnologie van andere bedrijven gebruiken, en anderzijds deze technologieën ook zelf (verder) ontwikkelen.

In de figuur staat langs de horizontale as de afstand tot de markt uit met spelers in de keten die dichter of verder van de markt afstaan. Verticaal zijn de volgende lagen te onderscheiden:

1. de basistaalvoorzieningen, uitdrukkelijk beperkt tot de data zelf en het aanmaakproces ervan;
2. TST-onderzoek en ontwikkeling uiteindelijk resulterend in beschikbare TST-componenten;
3. TST-embedding
4. de vraagzijde.

Lagen 1-3 vormen samen het aanbodsysteem.



Gelaagd model voor het TST-innovatiesysteem (figuur overgenomen uit de *Technologie-verkenning Nederlandstalige taal- en spraaktechnologie*, M&I/Partners BV en Montemore NV (2004))

<sup>1</sup> Deze tekst is (bijna volledig) overgenomen uit de STEVIN-programmatekst, zie [http://www.stevin-tst.org/documenten/stevin\\_mjp\\_150904.pdf](http://www.stevin-tst.org/documenten/stevin_mjp_150904.pdf), paragraaf 5.1.

### Laag 1. TST-basisvoorzieningen

TST-basistaalvoorzieningen zijn ofwel generisch van aard (voor een bepaalde taal) en veralgemenend ofwel zijn ze gericht op de verdere ontwikkeling van bepaalde toepassingen. Ze zijn essentieel zowel voor het onderzoek als voor een brede waaier van productontwikkelingen. Vanaf het begin moet herbruikbaarheid over een langere periode nagestreefd worden. Gezien het arbeidsintensieve karakter moet dit type basistaalvoorziening bij voorkeur slechts een keer (met gezamenlijke inspanning) vervaardigd worden en vervolgens breed beschikbaar gemaakt worden tegen een redelijke prijs. In het bijzonder betreft het vooral geannoteerde lexicons, thesauri en corpora.

### Laag 2. TST-onderzoek en ontwikkeling

Laag 2 heeft betrekking op het fundamentele en strategische onderzoek dat door kennisinstellingen en bedrijven wordt uitgevoerd en aan de ene kant noodzakelijk is voor de ontwikkeling van bepaalde basistaalvoorzieningen in laag 1 en aan de andere kant voor de ontwikkeling van de TST-componenten in laag 2. In deze laag horen thuis:

- regels en grammatica's zoals o.a. fonologische regels voor grafeem-foneem omzetting; grammatica's voor syntactische analyse; morfologische regels etc.
- modules en basiscomponenten zoals grafeem-foneemomzetting; tekstvoorverwerking (o.a. het detecteren van zinsgrenzen, datums, eigennamen, tijdstippen, afkortingen); morfologische analyse; syntactische analyse; semantische analyse; tekstgeneratie; spraaksynthese; spraakherkenning; foneem-grafeemomzetting; prosodiegeneratie; prosodieherkenning; sprekerherkenning; taal- en dialectidentificatie.

De componenten kunnen onderverdeeld worden in twee klassen:

a) modules zijn componenten die verdere ontwikkeling ondersteunen. De gebruiker is bij voorkeur een andere TST-expert die de zwaktes kan relativeren en weet te omzeilen. Modules zullen veelal puur ondersteunend werken in het ontwikkelingsproces van een andere TST-component. Anderzijds kunnen ze ook dienen als basistechnologie waaruit een commercieel product wordt ontwikkeld.

b) halffabrikaten zijn afgewerkte producten die klaar zijn voor integratie in andere halffabrikaten of rechtstreeks in eindproducten. De klant is mogelijk een andere TST-ontwikkelaar die dit halffabrikaat in een groter geheel inbouwt, maar wellicht frequenter is het een niet-TST-specialist die deze kant-en-klare module in zijn toepassing inbouwt.

In de overgrote meerderheid van de toepassingen vormt TST slechts een component(je) van een veel groter geheel en soms is de eindgebruiker zich weinig bewust van de onderliggende technologie. Voorbeelden zijn legio: spellingcheckers in wordprocessors; spraaksynthese die gebruikt wordt om informatie die is opgeslagen in een database weer te geven over de telefoon; spraakherkenning die een informatiesysteem of een boekingsstelsel stuurt.

### Laag 3. Applicatieontwikkeling (TST-embedding)

Laag 3 heeft betrekking op de toegepast onderzoek naar en ontwikkeling van toepassingen voor eindgebruikers waarin gebruik gemaakt wordt van een TST-component. Het profiel van de integratoren is zeer uiteenlopend. Het betreft onder meer: integratoren in de telefonie (Logica, VOXTRON) die werken in opdracht van de dienstensector (bijvoorbeeld banken, openbaar vervoer, telecom operatoren); ontwikkelaars van consumentendiensten en/of professionele apparatuur (Philips, Nokia, Bosch, Siemens); software ontwikkelaars (content management, tekstverwerking, vertaler *workbenches*, taalleerssoftware), etc.