

# Taal is emotie

---

Yola Park

De computer-taalkundige en schrijver Hugo Brandt Corstius zei het al, “Wat men ook doet, de semantiek gooit roet.” Dat is geen bemoedigende beginselverklaring voor wie semantische interoperabiliteit nastreven. De termen semantiek en interoperabiliteit zijn allebei, precies, semantisch beladen door een bonte variëteit interpretaties.

Met dit hoofdstuk als context bedoelen we met interoperabiliteit uitwisselbaarheid van gegevens tussen systemen. Onder semantiek verstaan we de betekenis van een woord of zin. Dan staat semantische interoperabiliteit voor betekenisvolle uitwisseling van gegevens tussen systemen. Het gaat dus niet over de manier waarop gegevens worden uitgewisseld, maar over wat de gegevens betekenen.

## **Betekenisvolle uitwisseling van gegevens**

Dit roept de vraag op wanneer gegevens betekenisvol zijn uitgewisseld. Iedereen die zich met semantische interoperabiliteit bezighoudt, erkent dat afspraken maken over de inhoud moeilijker is dan over de techniek. Er worden jaren besteed aan het ontwikkelen van instrumenten zoals classificaties, taxonomieën en uitwisselingsmodellen om die zogenaamde eenheid van taal te bereiken. Zulke ijver is weliswaar lovenswaardig, maar kansloos! Dergelijke instrumenten zijn slechts symptoombestrijders. Eventueel brengen ze enige verlichting bij begripsverwarring, maar ze doen geen recht aan de werkelijkheid.

We kunnen variaties in betekenis proberen terug te brengen tot één definitie en afbeeldingsregels bedenken om begrippen tot elkaar te herleiden. Vroeg of laat gaat het echter wringen. Mensen hebben nu eenmaal de ruimte nodig om een woord met de juiste lading uit te drukken. En niemand die dat kan tegenhouden, taal is immers van iedereen. Deze gevoelsinhoud laat zich niet simpelweg standaardiseren.

Is het streven naar semantische interoperabiliteit dan een hopeloze onderneming omdat systemen geen taalgevoel hebben? Gelukkig niet, dat zou een treurige en weinig creatieve conclusie zijn. Wel moet duidelijk zijn dat we toe zijn aan een paradigmaverschuiving. Daarin is een prominente rol weggelegd voor context, interpretatie en implicatie omdat deze verschijnselen mede de betekenis van gegevens bepalen.

## **Context**

Hetzelfde woord gebruiken ‘betekent’ niet per definitie hetzelfde bedoelen. De context helpt vaak bij het ontrafelen van gegevens.

In beperkte mate wordt nu al rekening gehouden met de context, meestal door specificatie van het subject of domein. Dat is een statisch gebruik van contextinformatie. Een andere beperking is dat de meeste domeinen multidisciplinair zijn en de bepaling wat een domein is daarom arbitrair.

Het belang van context betreft het interpreteren van gegevens. Niet voor niets staan in elke vorm van informatie-uitwisseling de contextvragen naar wie, wat, waar, wanneer en waarom centraal.

Er is nog geen beproefde methode om de variëteit in betekenis van gegevens op eenduidige manier via context uit te drukken. In deze 'context' verdient Meta-patroon voor stelselmatige betekenisordering van informatiekundig ontwerper Pieter Wisse onze aandacht. De wereld en onze taal verandert en dan heb je een modelleermethode nodig die zowel flexibel als stabiel genoeg is om de verschillen en overeenkomsten mee uit te drukken.

### **Interpretatie**

Wat zegt een koe? Boe! leren we onze kinderen. Maar de Engelsen verstaan Moo! En de Duitsers interpreteren het geluid als Muh! En de Fransen Meuh! Volgens de Sapir-Whorfhypothese wordt onze waarneming bepaald door de taal die we tot onze beschikking hebben.

Feitelijk gaat het niet om wat wordt bedoeld (even terzijde, zou de koe eigenlijk wat bedoelen met haar geloei?), maar om hoe de ontvanger deze gegevens interpreteert. Er zit vaak een vacuüm tussen enerzijds de verstrekker, anderzijds de ontvanger van gegevens. Beiden hebben hun eigen perceptie op de gegevens en vervolgens interpretatie ervan.

De uitwisseling van gegevens is meestal eenrichtingsverkeer van verzender naar verwerker van gegevens. Beter bewustzijn dat interpretatieverschillen bestaan en deze onderkennen als vorm van context, kan sterk bijdragen aan het detecteren van betekenisverschillen.

### **Implicatie**

Het ene woord heeft in een bepaalde context een andere implicatie dan in een andere. Er zijn gegevens waarvan het van cruciaal belang is dat ze goed worden geïnterpreteerd.

De beruchte Mars Climate Orbiter verongelukte door een misverstand over de eenheid. De gerecyclede navigatieprogrammatuur verwachtte gegevens in de krachteenheid pound, terwijl de Mars Climate Orbiter in newtons 'sprak.' Daardoor ging de ruimtesonde op de verkeerde hoogte de Mars-atmosfeer binnen en viel door de veel te hoge druk uit elkaar.[bron: NASA]

Je zult maar onder het mes liggen, wanneer de chirurg links als rechts interpreteert. De implicatie van een fietsbel die rechts in plaats van links is gemonteerd is verwaarloosbaar ten opzichte van dat abusievelijk geamputeerde been.

Bij het uitwisselen van gegevens behoren de gegevens waarvan de implicatie groot is het meeste aandacht te krijgen. Nu gebeurt doorgaans geen prioriteitstelling op basis van implicatie.

### **Het taalgevoel van computers**

Computers zijn rekenwonders, consequenter dan wie dan ook. Maar taalgevoel kan je ze niet toedichten. Ijskoud vergelijken ze gelijkvormige cijfers met elkaar zonder zich af te vragen of de uitkomst klopt volgens menselijke maat. Ook in deze context is het voorbeeld van de Mars Climate Orbiter illustratief.

We gebruiken taal om uit te drukken wat we denken en voelen, of althans wat we willen delen met anderen. We hechten eraan dat de woorden exact weergeven wat onze ziel beroert. Aan het taalvermogen van computers kunnen we weinig doen, dat is voorbehouden aan mensen. Aan de taalvaardigheid valt nog veel te verbeteren. Wat dacht u van een soort Nederlands als tweede taal (NT2) voor computers?

Uit de neurolinguïstiek weten we het één en ander over hoe hersenen taal begrijpen. Er is onderzocht welke woorden en zinnen (langer) actief worden gehouden in ons geheugen waardoor sneller andere woorden worden begrepen. De regels die we daaruit kunnen destilleren zouden we kunnen implementeren in informatiesystemen. Dan krijgen we voor gegevens een soort scenario-gebaseerde betekenistoekenning.

### Trippende informatiesystemen

Informatiesystemen zouden van het gebruik van LSD flink kunnen opknappen. We bedoelen hier met LSD natuurlijk het acroniem voor Luisteren, Samenvatten, Doorvragen.

Gegevens uitwisselen is maar een eenzame bezigheid. Systeem A levert gegevens aan systeem B voor verdere verwerking. Het zou al een stuk verbeteren als er wat meer inhoudelijke communicatie zou zijn tussen de systemen, zoals hierboven aanbevolen door verificatie van gegevens met hoge implicatie. Dus niet alleen technische verificatie of alle gegevens zijn overgekomen, maar tevens inhoudelijke verificatie of de betekenis juist is overgekomen.

We weten dat aannames een hoofdbron zijn van misverstanden. Dit speelt ook bij informatiesystemen parten. Wat is M/V volgens u? Ik veronderstel dat u hierbij denkt aan Man/Vrouw. In een andere context kan M/V ook staan voor Moeder/Vader. Bij inhoudelijke verificatie zouden dit soort verschillen kunnen worden ontdekt. Nou, laten we die inhoudelijke verificatie ook doen!

### Taal is emotie

Hopelijk zien steeds meer mensen de noodzaak tot een paradigmawissel. Zonder context, interpretatie en implicatie in acht te nemen en zonder inhoudelijke verificatie is semantische interoperabiliteit een fata morgana. Taal is emotie. Als we daaraan voorbijgaan, blijven we qua semantiek op een eiland zitten en wordt het nooit wat met die interoperabiliteit.

#### Yola Park

studeerde Taal en Informatica aan de Universiteit van Tilburg. Ze werkt bij de Koninklijke Bibliotheek als technisch projectmanager van het project Improving Access to Text (IMPACT). Doel van dit Europese project is om innovatieve oplossingen te vinden voor de verbetering van optische tekenherkenning (OCR) en methodes om het proces van massadigitalisering van gedrukte historische tekst in Europa te versnellen.