



Antal van den Bosch (1969) studeerde Taal- en literatuurwetenschappen aan Tilburg University en promoveerde aan Maastricht University. Sinds 2011 is hij hoogleraar Taal- en spraaktechnologie aan de Radboud Universiteit. In 2012 werd hij lid van de KNAW. Daarnaast is hij onder meer gasthoogleraar aan het Computational Linguistics and Psycholinguistics Research Centre van de Universiteit Antwerpen en verbonden aan het Netherlands eScience Center.

GEEN GRENS TUSSEN ALFA EN BÈTA

Antal van den Bosch trad op 1 januari 2017 aan als directeur van het Meertens Instituut. Hij blijft ook twee dagen in de week hoogleraar Taal- en spraaktechnologie aan de Radboud Universiteit. De stap naar het Meertens Instituut vindt hij min of meer logisch: 'De geesteswetenschappen zijn de natuurlijke habitat van computationeel linguïsten.'

Door Edith van Gameren

Foto Elodie Burrillon

Hoe ben je vanuit de informatica bij het Meertens Instituut voor Nederlandse taal en cultuur terechtgekomen?

'Er werken drie instituten samen binnen het KNAW humanities cluster, die samen een digital humanities groep wilden starten. Ze zochten iemand die daarin thuis was. Ik was al nauw betrokken bij het eerste programma op dit gebied, CATCH, en ik heb mijn wortels in de taalhoek. We willen de computer leren met natuurlijke taal om te gaan. Daarnaast biedt taaltechnologie gereedschappen voor andere wetenschappen. Het is daarom niet vreemd om computerlinguïsten in de geesteswetenschappen tegen te komen: historici en taalkundigen hebben allemaal teksten waar ze onderzoek op doen.'

Wat merk je van een scheidslijn tussen de vakgebieden?

'Als je in de beide vakgebieden actief bent, ervaar je geen grens, voor mij is het gewoon één gebied. Maar in Nederland schept de buitenwereld een soort schizofrene grens tussen alfa en bèta, en zijn er vooroordelen over en weer. Mensen uit de bètahoek zijn vaak onbekend met methoden en technieken van onderzoek in alfa- en gammawetenschappen. Ik heb vaak meegemaakt dat mensen gingen samenwerken en zich daarom voor het eerst in elkaars vakgebied moesten verdiepen. Dat levert voor beiden vaak interessante resultaten op.'

Wat gaat er veranderen bij het Meertens Instituut?

'Het Meertens Instituut is een van die bijzondere plekken waar taal- en cultuuruitingen in heel veel vormen bestudeerd worden. Waar het Meertens sterk in is - etnologie en taalkunde - moet sterk blijven. Big data en taaltechnologie kunnen dat ondersteunen. Sociolinguïsten doen onderzoek naar grote hoeveelheden data, ik kan daar mijn kennis over big data aan toevoegen. Twitter en andere sociale media zijn zulke ongelofelijk rijke bronnen waar je zo veel vragen op los kunt laten.'

Kun je een voorbeeld geven van zo'n vraag?

'Op Twitter bijvoorbeeld bestaan relaties tussen mensen; je kunt zien wie elkaar volgt. Als je daarop inzoomt en je zoomt nog verder in op relaties van mensen die elkaar over en weer noemen in tweets, dan heb je een sociale relatie te pakken. Binnen zo'n groepje is vaak sprake van een sociolect, een groepstaal. De data uit zo'n groep kun je gebruiken om taalmodellen te maken, bijvoorbeeld om een autocorrect of auto-completion functie te personaliseren. Van één persoon heb je vaak te weinig data, omdat veel mensen niet zo veel twitteren. Door het systeem te laten leren van de taal van vrienden, kun je als het ware een taaltransplantatie doen. Dat is met name voor mensen met een handicap, die heel veel behoefte hebben aan een goed werkende autocompletion functie, heel handig.'