

Hulstijn, J.H. (2003). Vacature Appel. In F. Kuiken (red.), *Appel heeft het gedaan* (pp. 172-178). Amsterdam: uitgeverij Bakker.

Schrijver en taalwetenschapper: Appel schrijft misdaadromans en “spannende” verhalen en houdt zich als taalwetenschapper onder andere bezig met de verwerving van het Nederlands als tweede taal door in Nederland wonende anderstaligen (hierna *NT2-ers*). Zou Appel, tijdens het schrijven van zijn romans en verhalen, aan NT2-ers denken als mogelijke lezersgroep, vroeg ik mij af. Bij NT2-ers hoeven we niet alleen te denken aan asielzoekers. Die leren het Nederlands immers niet met het doel om zich onmiddellijk op het werk van een Mulisch, Haasse of Appel te storten. Maar er zijn ook vele tienduizenden anderstaligen die goed ingeburgerd zijn, best wel nader kennis willen maken met wat de Nederlandse literatuur te bieden heeft en ook wel eens zin hebben in een spannend boek. Hoe toegankelijk is het werk van Appel voor díé NT2-ers? Die vraag wil ik in dit opstel, althans ten dele, proberen te beantwoorden. Ik beperk me tot een analyse van de woordenschat van één van Appels verhalen, gekozen uit de verhalenbundel “Met angst en beven”, in 2000 uitgegeven bij Ooievaar in Amsterdam. In dit verhaal, waarvan ik de titel nog even niet vermeld, spelen naar mijn oordeel typisch Nederlandse omstandigheden en feiten minder een rol dan in de andere verhalen uit de bundel (of in de romans van Appel), en dat maakt dit verhaal mogelijk meer geschikt als lectuur voor NT2-ers.

Ik begin met een simulatie van de situatie waarin de NT2-leerder verkeert die alle 2044 woorden kent uit het Basiswoordenboek Nederlands (De Klein & Nieuwborg, 1983), dat als standaard fungeert voor basisleergangen NT2 en voor NT2-examens. Ik noem onze denkbeeldige NT2-leerder Maxima. Laten we aannemen dat Maxima precies deze 2044 woorden van het Nederlands kent en geen andere. Dan zien de eerste twee alinea’s er als volgt uit, waarbij de sterretjes staan voor woorden die zij nog niet kent.

Acht uur. Vroeg genoeg om nog even de krant in te kijken. Ze liep naar beneden. De drie trappen naar boven hield ze de krant onder haar arm *1*. Een half jaar en langer geleden had ze vaak beneden al, ongemakkelijk op een trap*2* zittend, de *3* gepakt met de meeste personeels*4* en was ze de aangeboden *5* langsgegaan. Maar nu liet ze zich niet meer zo *6*. Haast was nergens voor nodig, zeker niet op zaterdagochtend.

Ze legde de krant op tafel, liet haar ochtendjas op de grond glijden en ging onder de *7* die ze drie jaar geleden in het keukentje had laten inbouwen. Frans had dat gedaan. Die was gelijk met haar afgestudeerd. Met zijn twee rechterhanden verdiende hij er zoveel bij in het zwarte *8*, dat hij er ruim van kon leven. Ze was hem uit het oog verloren, zoals bijna iedereen achter de *9* van haar dagelijks leven was verdwenen.

Het stukje bevat 155 woordtekens en negen ervan zijn Maxima onbekend; de overige 146 kent ze wel. Haar woordkennis zorgt dus voor een “tekstdekking” van 94% van de woordtekens (146/155). Ik heb Maxima’s kennis overigens wel enigszins opgerekt want ik ben ervan uitgegaan dat samengestelde woorden die

zelf niet in de 2044-basislijst voorkomen, maar waarvan de samenstellende delen er wel in voorkomen, wel enigszins begrijpelijk zullen zijn, zoals *ongemakkelijk, langsgedaan, zaterdagochtend, ochtendjas, inbouwen, afgestudeerd, rechterhanden*. Voordat ik verder ga met de lexicale analyse van het verhaal vermeld ik eerst even enkele begrippen en feiten uit de lexicale statistiek.

In de lexicale statistiek onderscheidt men woord*lemma*'s, woord*types* en woord*tekens*. Een lemma is een woord dat we in een woordenboek kunnen vinden, zoals *laten* (maar niet *liet* of *gelaten*), *krant* (maar niet *kranten*) en *de*. Een woordtype is een woordvorm. Er zijn meer types dan lemma's want *liet*, *gelaten*, en *laten* zijn woordvormen maar alleen *laten* is bovendien ook lemma; en *krant* en *kranten* zijn types maar alleen *krant* is een lemma. Een woordteken is een daadwerkelijk in een tekst voorkomend woordtype. Dikwijls komt een woordtype meer dan eens voor en we spreken dan van even zovele woordtekens. Van het woordtype *liet* bijvoorbeeld komen in de boven geciteerde eerste twee alinea's van "Vacature" twee tekens voor, er staan drie woordtekens van het type *krant* in en 10 woordtekens van het type *de*. Kortom, de tekens van een tekst behoren tot een kleiner aantal types dan er tekens zijn, en de types behoren op hun beurt weer tot een kleiner aantal lemma's.

In haar proefschrift "Een keur van woorden: De wenselijke en feitelijke receptieve woordenschat van anderstalige studenten" (Vrije Universiteit, 1994) geeft Suzanne Hazenberg voorbeelden van tekstdekkingspercentages. Ze laat zien hoeveel procent van het zogenaamde INL-tekstcorpus 'gedekt' wordt door hoeveel woorden. Het INL-corpus is een door het Instituut voor Nederlandse Lexicografie in Leiden aangelegd bestand van teksten. Ten tijde van Hazenbergs onderzoek omvatte het INL-corpus 930 integrale boeken, zowel fictie als nonfictie, uitgegeven tussen 1970 en 1988, samen goed voor ruim 42 miljoen woordtekens. Met de frequentie waarin woorden van een taal voorkomen, is het merkwaardig gesteld. Relatief weinig woorden komen heel veel voor en relatief veel woorden komen zelden voor. Zo zijn er vier woorden (namelijk *de*, *van*, *en* en *een*) die elk meer dan een miljoen keer voorkomen in het INL-corpus (samen 5.680.893 woordtekens, goed voor 13,4% tekstdekking); 375 woorden (lemma's) komen elk meer dan 10.000 keer voor (66% dekking); 2918 woorden komen meer dan 1000 keer voor (82%) en 11.123 woorden komen meer dan 100 keer voor (89%). Maar liefst 52 duizend woorden komen elk slechts één keer voor in het INL-corpus.

Wat betekenen deze cijfers voor NT2-leerders? Het goede nieuws is dat je met het leren van bijvoorbeeld de 2918 meest frequente woorden al zo'n 82% dekking van een gemiddelde Nederlandse tekst haalt. Het slechte nieuws is dat 82% tekstdekking *niet* 82% tekstbegrip inhoudt. Met 82% dekking begrijp je eigenlijk nog niets; je komt niet veel verder dan het vaststellen van genre en onderwerp van de tekst. Pas met 95% tekstdekking ga je een tekst begrijpen

zonder om de haverklap het woordenboek te hoeven raadplegen. Hazenberg trekt de conclusie dat frequentie voor NT2-leerders alleen een zinnige leidraad voor het leren van het Nederlands vormt als het gaat om de woorden die meer dan 100 keer voorkomen in het INL-corpus (11.123 lemma's), want die woorden komen voor in teksten over allerlei onderwerpen. (Deze lijst van 11.123 woorden staat in het wereldje van NT2-deskundigen bekend als de H&H-lijst, ofwel Hazenberg&Hulstijn-lijst.) Als je meer woorden wilt leren, moet je veel teksten lezen die je nuttig of interessant vindt en zelf bepalen of je de nog onbekende woorden zult overslaan dan wel in het woordenboek zult opzoeken en leren.

Na deze statistische uitweiding zijn we klaar voor een analyse van Appels verhaal. Het hele verhaal telt 7079 tekens (blz. 339-358), ofwel 1785 types. Maxima, met haar kennis van de 2044-basislijst, inclusief transparante samenstellingen en de in het verhaal voorkomende namen van personen en een poes (69 naamtekens), komt op 6095 bekende tekens, ofwel een tekstdekking van 86%. Daarmee begrijpt ze nog maar nauwelijks iets van de tekst. Dit verhaal is dus zeker geen lectuur voor beginnende NT2-leerders. Kent Maxima echter Engels of Frans, of een andere taal die veel leenwoorden uit die talen bevatten (zoals het Spaans!), dan gaat de tekstdekking flink omhoog. Van de negen ontbrekende woorden uit de boven weergegeven eerste twee alinea's (in volgorde: *geklemd*, *trede*, *katernen*, *advertenties*, *vacatures*, *opjagen*, *douche*, *circuit*, *horizon*) kunnen we er wel een stuk of vijf beschouwen als betrekkelijk transparante leenwoorden voor iemand die een Romaanse of Germaanse taal kent. Ik schat dat, als we zulke transparante leenwoorden meetellen, Maxima al gauw op 90% tekstdekking komt. Maar 90% is nog niet genoeg om een tekst zonder frustratiegevoelens te kunnen lezen en Maxima zal nog stevig door moeten leren om van Appels verhaal te kunnen genieten. Als ze het advies van Hazenberg opvolgt en de 11.123 woorden bevattende H&H-lijst leert, dan komt ze op het zeer comfortabele niveau van maar liefst 98%! Gaan we bovendien uit van de niet onredelijke veronderstelling dat ze het Engels goed beheerst (er komen enkele Engelse uitdrukkingen in de tekst voor die kennelijk op last van de purist Appel cursief moesten worden afgedrukt: *airmiles*, *senior consultants*, *young executives*, *double-breasted*, *The Economist*, *direct mail*, *fuck off*, *life goes on*), en nemen we ook aan dat ze inmiddels ook vertrouwd is geraakt met namen en begrippen zoals *NRC-Handelsblad*, *NOS*, *CDA*, *mavo*, *BMW*, *Rokin*, *Spui*, *Amsterdam*, *Zuid-Limburg*, *Vaals*, dan zit ze zelfs op 99%.

In de tekst staan nog 48 woordlemma's (54 woordtekens) die niet in de H&H-lijst voorkomen, geen Engels zijn en ook geen namen. Daaronder bevinden zich *aanrecht*, *armetierig*, *drab*, (*naar de*) *filistijnen*, *hartgrondig*, *heft*, *hufter*, *kaftje*, *ketsen*, *klauwde*, *plons*, *stekker*, en de spreektaalwoorden *sodemieter*, *klootzak*, *verdomd*, *verdomme* en *proost*. Dat laatste woord, *proost*,

is overigens essentieel voor een begrip van de pointe in de laatste zinnen van het verhaal:

Ze trok beide flessen open, zette twee glazen neer, en schonk ze beide vol. Ze pakte er één en hief het hoog in de lucht. ‘Proost, Leo.’

Concluderend kunnen we zeggen dat Appels misdaadverhaal in lexicaal opzicht nog te lastig is voor een NT2-leerder op beginnersniveau (met een woordenschat van circa 2000 woorden). Maar als zij of hij een woordenschat verworven heeft van zo’n 10.000 woorden, dan kunnen we dit spannende Appel-verhaal zeker aanbevelen. Overigens is 10.000 nog bescheiden als je bedenkt dat Van Dale’s handwoordenboek ruim 50.000 ingangen telt en de driedelige grote Van Dale zo’n 240.000.

Het wordt nu tijd om de titel van het verhaal te vermelden. Het draagt draagt de veelzeggende titel “Vacature”, een woord dat niet in de 2044-lijst voorkomt. Als Maxima, toen haar woordenschat zich nog beperkte tot precies die woordenlijst, dat woord in het woordenboek had opgezocht, dan was het begin van het verhaal minder onbegrijpelijk geweest.

Rest mij de hoop uit te spreken dat het er bij het vervullen van de *vacature Appel* die door het vertrek van René Appel ontstaat aan de Faculteit Geesteswetenschappen van de Universiteit van Amsterdam, minder misdadig zal toegaan dan in *Appels Vacature*. Ter voorbereiding op mijn rol bij de selectie trek ik twee flessen open, zet twee glazen neer, schenk ze beide vol en hef er één hoog in de lucht. ‘Proost, René.’¹

¹ Met dank aan Bregje van Oel voor het scannen van “Vacature” en aan Arjen Florijn voor het maken van een woordfrequentielijst.