

Voldoen vereenvoudigde romans aan voorwaarden voor NT2-woordenschatverwerving?

JET DOEDEL

Als volwassen NT2'ers veel en vaak lezen, kan hen dat helpen om nieuwe woorden in de tweede taal te leren. Tenminste, als daarvoor geschikt leesmateriaal wordt gebruikt. Geschikt betekent in dit geval: met voldoende bekende woorden om de tekst te kunnen begrijpen, met een aantal onbekende woorden om te kunnen leren, en zorgend voor genoeg leesplezier om te blijven lezen. Voor dit doel lijken hertalingen — (literaire) romans die voor volwassen taalleerders zijn geschreven in eenvoudige taal — een goede optie. Maar voldoen de Nederlandse hertalingen die bedoeld zijn voor ERK-niveaus A2 en B1 aan de voorwaarden voor woordenschatverwerving door extensief lezen? Om die vraag te beantwoorden, zijn in dit onderzoek 54 hertalingen langs de lat van de vijfduizend meest-frequente woordfamilies in het Nederlands gelegd. Bij 52 van de 54 boeken bleek de dekkingsgraad hoger dan 95 procent — een minimumdrempel voor tekstbegrip —, met een gemiddelde dekkingsgraad van 97 procent. Alle boeken boden voldoende blootstelling aan een aantal ongedekte woorden om kennis te kunnen verwerven over hun vorm-betekenisrelatie.

Om kennis over woorden te vergaren, is het nodig dat een leerder een woord vaak tegenkomt. Dat kan in de context van het NT2-onderwijs op allerlei manieren gebeuren, variërend van intentionele, directe instructie en oefening tot incidentele, inhoudsgerichte ervaringen. Bij die laatste vorm, incidenteel leren, is het opdoen van taalkennis niet het primaire doel, maar een bijproduct van een andere activiteit. Voor moedertaalleerders wordt incidenteel leren gezien als de belangrijkste manier voor het vergaren van woordkennis (Nagy et al., 1985; Nation, 2001; Schmitt, 2000; Webb & Nation, 2017), maar ook volwassen tweedetaalleerders kunnen woordkennis incidenteel verwerven (Elgort et al., 2018; Feng & Webb, 2020; Grabe & Stoller, 2011; Horst et al., 1998; Pellicer-Sánchez, 2016). Volgens Webb en Nation (2017) geldt voor hen wel een belangrijk verschil: waar incidenteel woordleren bij moedertaalsprekers succesvol verloopt door de grote hoeveelheid input waaraan zij van nature blootstaan, is die input bij tweedetaalleerders vaak beperkter. Maar gezien het succes van incidenteel leren bij moedertaal-

Voldoen vereenvoudigde romans aan voorwaarden voor NT2-woordenschatverwerving?

leerders mogen de mogelijkheden ervan bij tweedetaalleerders niet genegeerd worden: ook hun woordenschatverwerving is gebaat bij gestage blootstelling aan overvloedige input (Day & Bamford, 1998; Grabe, 2009; Webb & Nation, 2017).

Ontmoetingen met woorden

Overvloedige input is belangrijk om woordkennis op te doen, maar hoe vaak moet een tweedetaalleerder een woord tegenkomen om het daadwerkelijk te kennen? Hierbij spelen individuele lezerskenmerken en de beoogde soort kennis een grote rol. In de literatuur wordt een waaier aan benodigde aantallen blootstellingen genoemd. Een veelgebruikte vuistregel bij lezen is dat acht tot tien ontmoetingen voldoende zijn om kennis op te doen over de vorm-betekenisrelatie van een woord (Schmitt, 2008), of om nieuwe woorden even snel en accuraat te lezen als reeds bekende woorden (Pellicer-Sánchez, 2016). Naast het benodigde aantal blootstellingen is ook relevant welke woorden leerders het beste kunnen leren. Dat zijn de woorden die het vaakst in het taalgebruik voorkomen (Gardner, 2013; González-Fernández & Schmitt, 2017; Nation, 2001, 2013, 2014; Schmitt & Schmitt, 2014; Webb & Nation, 2017). Hoewel de individuele woordenschat van taalleerders zich niet laat definiëren door een algemene woordenlijst, bestaan er in de praktijk richtlijnen voor het aantal meest-frequente lemma's dat minimaal receptief gekend moet worden om de ERK-niveaus te behalen: een NT2-leerder zou op A1-niveau ongeveer duizend woorden moeten kennen, op A2-niveau ongeveer tweeduizend, en op B1-niveau ongeveer vijfduizend (Bossers, 2022; Dolmans, 2015). Lesmethodes voor NT2 beogen naar die aantallen toe te werken (Bossers, 2022). Ook in het Staatsexamen NT2 wordt gewerkt met de ERK-niveaus

en daarmee met de bijbehorende aantallen woorden, waarbij de beheersingscriteria voor programma I zijn te vergelijken met de vaardigheden van ERK-niveau B1 (College voor Toetsen en Examens, z. d.)

Extensief lezen

Om te zorgen voor de overvloedige input die incidentele woordenschatverwerving faciliteert, kan extensief lezen helpen: het onafhankelijk en stil lezen van flinke hoeveelheden tekst (Feng & Webb, 2020; Horst, 2005; Nation & Waring, 2020; Pigada & Schmitt, 2006; Subramaniam & Zainal, 2016; Suk, 2017; Webb & Nation, 2017). Het lezen van langere teksten zorgt ervoor dat lezers woorden ontmoeten ingebed in een context, en zij gebruiken die context om woordbetekenis af te leiden (Bolger et al., 2008; Elgort & Warren, 2015; Grabe & Stoller, 2011; Feng & Webb, 2020; Nation, 2001, 2007; Webb, 2008). Om gebruik te kunnen maken van die context, is het nodig dat deze begrepen wordt: 'A word may not be learned even after many encounters if a reader's vocabulary is insufficient and if the text is not well understood' (Elgort et al., 2018, p. 342). Een goed ontworpen programma voor extensief lezen kan een zinvol onderdeel vormen van het NT2-curriculum voor volwassenen (González-Fernández & Schmitt, 2017; Suk, 2017), bij voorkeur in combinatie met intentioneel woordleren (Hu & Nation, 2000; Waring & Nation, 2004), om zo de kansen op woordleren te maximaliseren. Hierbij is een belangrijke rol weggelegd voor leesplezier, omdat plezier een sterke stimulans is voor leesmotivatie (Stokmans & Wolters, 2019). Als lezen plezierig is, bijvoorbeeld dankzij een goed romanverhaal, wordt het immers gemakkelijker om veel leesmateriaal te verwerken (Bruner, 1986). Voor extensief lezen is leesmateriaal nodig dat past bij het vaardigheidsniveau van de leerder,

ofwel: leesmateriaal dat gemakkelijk genoeg is om zonder significante leerlast gelezen te worden (Suk, 2017). Hier verschijnen hertalingen ten tonele: romans die qua vorm en inhoud zo zijn bewerkt dat ze door volwassen taalleerders op A2/B1-niveau gelezen kunnen worden (moedertaalsprekers die moeite hebben met lezen (NT1) behoren overigens ook tot de doelgroep van deze boeken).

Het gebruik van vereenvoudigde tekst voor tweedetaalleerders is niet zonder kritiek. Volgens critici worden structuur-, cohesie- en verbindings-elementen in eenvoudige tekst vaak achterwege gelaten, wat het tekstbegrip bemoeilijkt (Land, 2009; Phillips & Shettlesworth, 1988; Van Silfhout, 2014), en kunnen vereenvoudigde teksten door de veranderde woordenschat en syntaxis juist moeilijker te begrijpen zijn dan authentieke teksten (Long & Ross, 1993; Widdowson, 1978). Ook luidt de kritiek dat vereenvoudigde teksten tweedetaalleerders niet goed voorbereiden op en kennis laten maken met natuurlijke taal (Goodman & Freeman, 1993; Honeyfield, 1977; Larsen-Freeman, 2002). Anderen concluderen echter dat tekstbegrip bij vereenvoudigde teksten voor leerders op de begin- en middenniveaus juist hoger uitvalt dan bij niet-vereenvoudigde teksten (Crossley et al., 2014; Javourey-Drevet et al., 2022; Rets & Rogaten, 2021; Tweissi, 1998; Yano et al., 1994), en dat vereenvoudigde teksten juist meer coreferentiële samenhang en connectieven bevatten (Crossley et al., 2007). Daarnaast kan de woordenschat in niet-vereenvoudigde leesteksten, zoals kranten of romans, voor leerders te moeilijk zijn om te dienen als begrijpelijke input (McLaughlin, 1987; Nation, 2009; Schmitt, 2008; Shook, 1997). Omdat de woordenschatomvang van een lezer een belangrijke factor is voor de mate waarin deze een tekst begrijpt (Nation, 2006; Laufer & Ravenhorst-Kalovski, 2010), moeten teksten passen bij diens taalniveau (Suk, 2017; Webb & Nation, 2017). Volgens

Nation (2001) zijn teksten op het juiste niveau daarom een belangrijk middel voor taalleren: ‘Without simplification, [...] meaning-focused input [...] becomes impossible for all except advanced learners’ (p. 172). Day en Bamford (1998) maken in dit opzicht een vergelijking met literatuur die specifiek wordt geschreven voor kinderen of jongeren — zo is volgens hen ook voor tweedetaalleerders specifiek leesmateriaal nodig dat aansluit bij hun taalniveau.

Nation (2001) stelt dat de kritiek op de vereenvoudiging van tekst in feite vaak kritiek is op niet goed uitgevoerde vereenvoudiging. Voor dat vereenvoudigen bestaan allerlei richtlijnen (zie Inclusion Europe, 2009; Raad van Europa, 2022; Stichting Makkelijk Lezen, z. d.), bijvoorbeeld dat tekst op A2- en B1-niveau moet bestaan uit korte zinnen met gangbare woordvolgordes, die zijn geschreven in veelvoorkomende, alledaagse taal. Eventuele minderfrequente woorden kunnen in de lopende tekst of in een verklarende woordenlijst worden toegelicht. Voor een directe stijl en duidelijke verhaallijnen worden zo nodig ook de complexiteit van verhaallijnen, perspectieven en chronologie vereenvoudigd. Dergelijke aanpassingen moeten waarborgen dat een verhalende tekst goed leesbaar is voor ‘minder goede lezers’ (NT1) en ‘mensen die het Nederlands niet als eerste taal hebben’ (NT2) (Stichting Makkelijk Lezen, z. d.) Nation en Wang (1999) benadrukken bovendien dat vereenvoudigde boeken aantrekkelijk moeten zijn: ‘[Graded readers] should be readable, interesting and well-written. Just because the vocabulary and grammar are controlled, the quality of the writing should not suffer. A graded reader should tell a good story well’ (p. 375).

Lexicale dekking

Hoewel tekstbegrip van veel meer factoren afhangt dan het aantal bekende woorden in de tekst, zoals het gebruik van leesstrategieën en

de cognitieve vermogens, executieve functies en achtergrondkennis van de lezer (Cohen & Macaro, 2007; Elgort & Warren, 2015; Laufer, 1989), is woordenschatomvang wel een essentiële voorwaarde voor tekstbegrip (Song & Reynolds, 2022): om een tekst überhaupt te begrijpen, moet een lezer de meeste woorden in die tekst kennen. Bovendien kan een lezer zonder tekstbegrip geen betekenis van onbekende woorden uit de context afleiden (Bolger et al., 2008; Feng & Webb, 2020; Nation, 2001). Het aantal bekende woorden in een tekst wordt lexicale dekking genoemd en uitgedrukt in een percentage. In de literatuur worden verschillende percentages genoemd voor de dekkinggraad die minimaal nodig is voor tekstbegrip. Volgens Laufer (1989) begrijpen lezers de informatie in een tekst vanaf een lexicale dekking van 95 procent significant beter dan bij een lagere dekking, en volgens Liu en Nation (1985) kunnen lezers bij 95 procent dekking woordbetekenis raden uit de context. Ook wordt 97 procent genoemd voor ‘*very comfortable reading*’ (Nation & Wang, 1999, p. 361), en 98 procent voor het begrip van academische teksten (Schmitt et al., 2011), een bevredigend begrip van verhalende teksten in de tweede taal (Song & Reynolds, 2022) en om fictie te lezen zonder hulpmiddelen (Hu & Nation, 2000). De percentages variëren dus naargelang het beoogde begripsniveau, de tekstsoort en de mate waarin hulpmiddelen gebruikt kunnen worden, maar er bestaat brede consensus dat tekstbegrip hapert als de dekking zakt onder de grens van 95 procent. Nation (2013, p. 356) adviseert leerders en docenten daarom ‘[to] make sure that between 95-98 per cent of the running words in a chosen reader are already known’.

Vraagstelling

In dit onderzoek is nagegaan of Nederlandse hertalingen voldoen aan de voorwaarden

voor woordkennisverwerving via extensief lezen door A2/B1-lezers. Ofwel: zijn de hertalingen geschreven met gebruik van de hoogfrequente woorden die horen bij die niveaus, en komen woorden die daarbuiten vallen vaak genoeg voor om te leren over de vormbetekenisrelatie ervan?

Hertalingen

Om de lexicale dekkinggraad en ruimte voor woordleren in Nederlands vereenvoudigd leesmateriaal te meten, zijn 54 Nederlandse hertaalde romans langs de lat van de vijfduizend meestfrequente Nederlandse woordfamilies gelegd. De gebruikte romans zijn verschenen bij uitgeverij Eenvoudig Communiceren, die de boeken zelf inschaalt op A2/B1-niveau, en voorzien van het keurmerk Makkelijk Lezen van de gelijknamige stichting. Dat keurmerk wordt verleend als een boek voldoet aan eisen aan de verhaal- en zinslengte, de inhoud, verhaal- en zinsstructuur, woordenschat en opmaak.

De 54 geanalyseerde romans zijn hertaald door verschillende hertalers (waaronder de auteur van dit artikel) en bedoeld voor een volwassen publiek. Om de boeken geschikt te maken voor de dekkinggraadanalyse is alle informatie van en over de uitgever verwijderd, evenals colofons en kop- en voetteksten; behouden zijn inhoudsopgaven, opdrachten, voorwoorden en woordenlijsten. Ook zijn eigennamen (waaronder aardrijkskundige namen) en hun afleidingen uit de romans verwijderd. Eigennamen staan immers los van de eerdere woordenschatverwerving, kunnen goed uit de context begrepen worden en zijn door hun initiële hoofdletter goed herkenbaar als eigenaam (Hirsh & Nation, 1992; Nation & Wang, 1999). Ook niet-spellingseigen accenten (écht, ík) en getallen in getalnotatie zijn verwijderd uit de tekst. Deze aanpassingen leverden een corpus op van in totaal 869.107 tokens of 18.804 types, wat neerkomt op

gemiddeld 16.095 tokens of 2.032 types per boek.

Referentielijst

Om de dekkinggraad in de hertalingen te meten, is een referentielijst nodig. Hiervoor is in dit onderzoek *A frequency dictionary of Dutch* (Tiberius & Schoonheim, 2014) als basis gebruikt, een woordenboek met de vijfduizend meestfrequente Nederlandse lemma's. Omdat romantiek echter niet enkel bestaat uit lemma's maar ook uit allerlei vervoegingen en vervormingen daarvan, is deze lemmalijst met regelmatige, veelvoorkomende vormen uitgebreid naar woordfamilies, analoog aan de aanpak van Nation en Wang (1999) in hun onderzoek naar de dekkinggraad in Engelse *graded readers*. Als een lezer bekend is met de basisvorm van een woord en affixen, mogen deze immers als bekend verondersteld worden (Hirsh & Nation, 1992). Op deze manier zijn bij werkwoorden onder meer vervoegingen (loopt) en bepaalde verzelfstandigingen (inbreker) toegevoegd, bij zelfstandige naamwoorden meervouden (boeken), verkleiningen (boekje), sommige vrouwelijke vormen (danse-res) en bepaalde adjectiveringingen (mistig). Bij bijvoeglijke naamwoorden zijn affixen (leuke), trappen van vergelijking (mooier) en bepaalde verzelfstandigingen (droogte) toegevoegd. Ook verbuigingen en verzelfstandiging van telwoorden (tweede, honderden) werden opgenomen in de woordfamilies.

De originele lemmalijst bevatte al een aantal dubbele woorden (bijvoorbeeld *vers* als zelfstandig naamwoord en als bijvoeglijk naamwoord). Door het uitbreiden naar woordfamilies verschenen nog meer dubbelingen in de lijst (bijvoorbeeld *bedienden* als meervoudig zelfstandig naamwoord en als vervoegd werkwoord). Omdat dubbelingen de dekkinggraad kunnen vertroebelen en dus niet wenselijk zijn, is de lijst ontdebeld. Bij dubbelingen in de originele lemmalijst is steeds het woord met de hoogste

frequentie behouden. Bij dubbelingen die ontstonden door uitbreiding naar woordfamilies is steeds uitgegaan van hun positie in de lijst en semantische gerelateerdheid. Zowel bij semantisch niet-gerelateerde woorden (zoals *staart* als zelfstandig naamwoord en als werkwoord) als bij semantisch gerelateerde woorden (zoals *vraag* als zelfstandig naamwoord en als werkwoord) is telkens het hoogstfrequente woord behouden. Hoewel het te verantwoorden zou zijn om semantisch gerelateerde woordfamilies helemaal samen te voegen op de hoogstfrequente plek, is daarvoor niet gekozen om de uiteindelijke dekkinggraad niet te flatteren (zo werd *vraag* als hoogfrequenter zelfstandig naamwoord behouden, maar bleven *gevraagd* en *vragend* op de plek van het laagfrequenter werkwoord). Het ontdebelen van de lijst vroeg soms om het maken van wat arbitraire keuzes, maar die zijn steeds behoudend en zo consciëntieus mogelijk gemaakt om de betrouwbaarheid van de referentielijst te bewaken.

Om te kunnen bepalen bij hoeveel woorden de minimale dekkinggraad bereikt wordt, is de uitgebreide en ontdebeldde woordenlijst vervolgens verdeeld in vijf banden van oplopende frequentie. De dekkinggraad is gemeten met het tekstanalyseprogramma AntWordProfiler 4.2.0 (Anthony, 2023). Om de woorden die niet door de referentielijst werden gedekt te analyseren, zijn deze eerst gelemmatiseerd met het tekstanalyseprogramma Frog (Van der Sloot et al., 2018) en vervolgens is gekeken hoe vaak ze in de boeken voorkwamen. Ten eerste is gekeken naar ongedekte woorden die minimaal acht keer per boek voorkwamen, de vuistregel voor het aantal blootstellingen dat nodig is voor kennisverwerving over de vorm-betekenisrelatie van een woord. Maar omdat in de literatuur ook woordleren bij minder blootstellingen wordt beschreven, is ook gekeken naar ongedekte woorden die vijf tot en met zeven keer per boek voorkwamen. Tot slot

zijn met AntWordProfiler (Anthony, 2023) de delen van ongedekte samenstellingen en de verklarende woordenlijsten in een aantal hertalingen vergeleken met de woorden in de referentielijst.

Resultaten

Afgezet tegen de referentielijst met de vijfduizend meestfrequente Nederlandse woordfamilies bleken de 54 hertalingen een gemiddelde dekkinggraad van 97 procent te hebben. De dekkinggraad van 52 boeken was minimaal 95 procent, waarvan bij vijf boeken minimaal 98 procent. Er was variatie in de frequentieband waarin de minimale dekking van 95 procent werd bereikt: bij vijf boeken was dat met drieduizend woorden, bij 22 boeken met vierduizend woorden en bij 25 boeken met vijfduizend woorden.

Vervolgens is gekeken naar de woorden die wel voorkwamen in de hertalingen, maar niet in de referentielijst, ofwel de woorden die uitgaand van de vijfduizend meestfrequente woordfamilies als onbekend verondersteld werden. Uitgedrukt in lemma's bevatte het gehele corpus 12.513 woorden, waarvan gemiddeld 263 lemma's per boek niet gedekt werden door de referentielijst. Van deze ongedekte lemma's kwam het overgrote deel slechts een of enkele keren in een boek voor. Maar alle boeken bleken ook ongedekte woorden te bevatten die zorgden voor voldoende blootstelling voor kennisverwerving over de vorm-betekenisrelatie. In totaal kwamen 717 ongedekte woorden minimaal vijf keer in een boek voor. Daarbij ging het gemiddeld om negen woorden die vijf tot en met zeven keer per boek voorkwamen, met een gemiddelde frequentie van zes keer (bijvoorbeeld *servet*, *coupé*), en gemiddeld acht woorden die minimaal acht keer per boek voorkwamen, met een gemiddelde frequentie van vijftien (bijvoorbeeld *trommel*, *herberg*).

Vervolgens is gekeken naar de aard van deze 717 ongedekte woorden. Deze bleken voor 61 procent te bestaan uit ongelede woorden (*biljart*, *divan*) en voor 39 procent uit samenstellingen (*boekenkast*, *halsketting*), die op hun beurt weer bestonden uit 373 unieke delen (in dit geval *boek*, *kast*, *hals* en *ketting*). Omdat kennis van die delen ook (enige) kennis over de samenstellingen kan impliceren, is gekeken of deze delen ook zelfstandig in de referentielijst voorkwamen. Dat was bij 83 procent van de delen van de samenstellingen het geval. Bijvoorbeeld: *fruitboom* staat niet in de referentielijst en werd dus als onbekend verondersteld, maar *fruit* en *boom* staan wel in de lijst, waardoor de lezer toch ook kennis over *fruitboom* kan hebben. Er waren echter ook samenstellingen waarvan de delen dan wel zelfstandig in de referentielijst staan, maar waarvan de betekenis niet transparant is. Bijvoorbeeld: *avond* en *klok* staan wel in de lijst, maar de betekenis van *avondklok* is niet af te leiden uit de betekenis van de delen. Van de ongedekte samenstellingen die minimaal vijf keer per boek voorkwamen en waarvan de delen in de referentielijst stonden, bleek 56 procent een transparante betekenis te hebben, wat kan impliceren dat de dekkinggraad wat hoger uitvalt dan gemeten.

Ook is gekeken naar de verklarende woordenlijsten achter in twintig hertalingen, waarin onbekend veronderstelde woorden uit de lopende tekst zijn opgenomen (bijvoorbeeld *barak*, *vonnis*). Deze woorden behoren wat betreft tekstbegrip en dekkinggraad tot een grijs gebied. Weliswaar moet de lezer het lezen onderbreken om de betekenis op te zoeken, maar deze drempel is relatief laag aangezien de woordenlijst binnen handbereik is, geschreven is in niveau-adequate taal, en past bij de context van het boek. In de verklarende woordenlijsten werden in totaal 384 woorden uitgelegd (range 4–61), waarvan 62 procent niet voorkwam in de referentielijst. Dit zijn dus ongedekte woor-

den waarover een lezer bij het opzoeken woordkennis kan opdoen. Tot slot worden in de hertalingen ook woorden uitgelegd in de lopende tekst. Een voorbeeld daarvan uit 't *Hooge Nest* (Van Iperen, 2022, p. 10): 'Ze moeten hun persoonsbewijs laten zien. Dat is een soort paspoort, met daarin hun naam, adres, pasfoto en vingerafdruk.' Dergelijke uitleg kan bijdragen aan tekstbegrip en woordkennis, al is dat niet terug te zien in de lexicale dekkingsgraad.

Conclusies

Voldoen Nederlandse hertaalde romans voor A2/B1-lezers aan de voorwaarden die NT2-woordenschatverwerving mogelijk maken? Met andere woorden: zijn ze geschreven in taal die tekstbegrip faciliteert, en bieden ze ruimte om nieuwe woorden te leren doordat onbekend veronderstelde woorden vaak genoeg voorkomen? Om deze vragen te beantwoorden zijn de dekkingsgraad en de frequentie van ongedekte woorden van 54 Nederlandse hertalingen geanalyseerd aan de hand van de vijfduizend meestfrequente woordfamilies. Veruit de meeste boeken (52 van de 54) bleken een dekkingsgraad van minstens 95 procent te hebben — waarvan vijf een dekkingsgraad van 98 procent —, met een gemiddelde dekkingsgraad van 97 procent. Hiermee voldoet het overgrote deel van de boeken aan het advies om te voorzien in minimaal 95 procent tekstdekking. Mogelijk valt de dekkingsgraad iets hoger uit, doordat lezers toch kennis kunnen hebben van ongedekte, onbekend veronderstelde samenstellingen (*fruitboom*) waarvan de delen (*fruit* en *boom*) wel bekend zijn. Ook de verklarende woordenlijsten achter in twintig van de 54 boeken en uitleg van woorden in de lopende tekst kunnen helpen om woordkennis op te doen, waardoor het tekstbegrip hoger kan uitvallen dan de gemeten dekkingsgraad.

Bij vijf boeken werd de minimale dekkingsgraad van 95 procent bereikt in de derde frequentieband, bij 22 boeken in de vierde frequentieband en bij 25 boeken in de vijfde frequentieband. Dit impliceert dat tekstbegrip op basis van woordenschat mogelijk is als lezers in grote lijnen respectievelijk de eerste drie-, vier- en vijfduizend meestfrequente woorden in het Nederlands kennen. Volgens de ERK-richtlijnen voor woordenschatomvang (Bossers, 2022; Dolmans, 2015) gaat het bij drie- en vierduizend woorden om het leesniveau van A2 naar B1 en bij vijfduizend woorden om niveau B1. De onderzochte boeken lijken dus wat betreft de gebruikte woordenschat te voldoen aan de voorwaarden die tekstbegrip op die niveaus mogelijk maken.

Daarnaast bleken de boeken gemiddeld negen lemma's per boek te bevatten die vijf tot en met zeven keer voorkwamen, en gemiddeld acht lemma's die acht keer of vaker voorkwamen — aantallen die genoeg zouden zijn om kennis over de vorm-betekenisrelatie van onbekende woorden op te doen. In de hertalingen lijkt dus ruimte te zijn om nieuwe woorden te leren. Daarbij geldt wel dat er heel wat gelezen moet worden om voldoende blootstelling te bereiken, terwijl een grote leeslast een drempel kan zijn voor minder taalvaardige of minder gemotiveerde lezers. Aandacht voor de juiste boekenkeuze, met een hoge dekkingsgraad en mogelijkheden voor leesplezier, is dus geboden.

Discussie

Op basis van de dekkingsgraad en de frequentie van ongedekte woorden lijken de hertalingen zich dus te lenen voor NT2-woordenschatverwerving, maar in de smalle focus van dit onderzoek zijn wel wat slagen om de arm op hun plaats. Er is gewerkt met dekkingsgraadpercentages, frequentie-

woordenlijsten, ERK-richtlijnen voor woordenschatomvang en blootstellingsaantallen voor woordkennisverwerving. Maar op elk van die parameters is aan te merken dat ze weliswaar gestoeld zijn op ruim onderzoek, maar geen onwrikbare criteria vertegenwoordigen. Zo vormt de minimale woorddekking van 95 procent een indicatie, maar geen garantie dat een lezer een tekst begrijpt. Door allerlei individuele kenmerken (cognitief vermogen, leesvaardigheid in de moedertaal, achtergrondkennis) kan een tekst zelfs bij een hogere dekkingsgraad te moeilijk zijn. Bovendien vereist tekstbegrip niet alleen woordkennis, maar ook andere (taal)vaardigheden (Hu & Nation, 2000). Hier staat tegenover dat lezerskenmerken en tekstuele eigenschappen (inhoud, grammaticale aanwijzingen, illustraties) er ook voor kunnen zorgen dat het tekstbegrip juist gunstiger uitpakt dan de lexicale dekkingsgraad doet vermoeden (Laufer, 1989). Bovendien is er in de NT2-lescontext geen bezwaar om hulpmiddelen in te zetten: zeker het gebruik van een woordenboek kan het leerrendement vergroten (zie Webb & Nation, 2017).

Ten tweede is de woordenschat van een individuele taalleerder niet zomaar te vangen in een afgebakende woordenlijst of ERK-niveau. Misschien kent een leerder minder of juist meer woorden, en ook de diepte van de woordkennis is individueel en lastig peilbaar. Referentielijsten vormen een geformaliseerde indicatie van wat iemand op een bepaald niveau beheerst, maar ze weerspiegelen nooit de eigenschappen van individuele lezers, terwijl die wel sterk van invloed kunnen zijn op het tekstbegrip. En tot slot is er dan wel consensus over het aantal blootstellingen dat nodig is voor woordkennisverwerving, maar is ook dit geen hard criterium. In de praktijk spelen de cognitieve eigenschappen van de lezer en de leercontext een grote rol (Hiebert & Kamil, 2005).

Allerlei lezers- en tekstkenmerken zorgen er dus voor dat de conclusies in dit onderzoek niet onweerlegbaar, maar wel richtinggevend zijn. Binnen de genoemde nuanceringen bieden ze een eerste inzicht in de vraag of de hertalingen überhaupt geschikt kunnen zijn voor woordenschatverwerving door extensief lezen op A2/B1-niveau. Dat lijkt het geval: veruit de meeste boeken zijn geschreven in woorden die passen bij de beoogde niveaus, en in elk boek komt een aantal ongedekte woorden vaak genoeg voor om vorm-betekenis-kennis op te kunnen doen. Daarmee lijken de boeken zich te lenen voor gebruik in extensieve NT2-leesprogramma's. De rol en ervaring van leesplezier bij hertalingen en lezersstrategieën bij de ontmoeting met onbekende woorden zouden het onderzoek kunnen zijn van verder onderzoek. In aanvulling op Nation (2001), Pellicer-Sánchez (2016), Pigada en Schmitt (2006), Suk (2017), Waring en Nation (2004), Webb en Chang (2015) en Subramaniam en Zainal (2016), die het positieve effect van het lezen van eenvoudige teksten op de woordenschatverwerving beschrijven, kan het daarnaast zinvol zijn om daadwerkelijke woordenschatverwerving door extensief lezen van Nederlandse boeken te meten en om deze te vergelijken met woordenschatverwerving door het lezen van niet-vereenvoudigde teksten. Dergelijk vervolgonderzoek kan laten zien of de boeken die qua woordgebruik en het percentage onbekend veronderstelde woorden geschikt lijken voor woordenschatverwerving daar inderdaad toe leiden, en hoe dit zich verhoudt tot woordenschatverwerving door het lezen van originele teksten.

In de praktijk

Wat betekent dit alles voor het gebruik van hertalingen in de context van de NT2-les voor volwassenen? Als de condities ide-

aal zijn (dus voldoende dekking voor tekstbegrip, ruimte voor woordleren en leesplezier), lijken deze geschikt materiaal te vormen voor woordenschatverwerving door extensief lezen. Ze kunnen bijdragen aan de omvangrijke input die NT2-leerders nodig hebben om hun woordenschat te verbreden, te verdiepen en te consolideren. Zoals ook González-Fernández en Schmitt (2017), Grabe en Stoller (2011) en Suk (2017) adviseren, kan extensief lezen daarom een zinvolle bijdrage zijn aan tweedetaalcurricula. In de praktijk zijn daarvoor drempels en hobbels te bedenken — tijd(gebrek), bestaande lesprogramma's en organisatie bijvoorbeeld. Toch kunnen extensieve leesprogramma's een welkom onderdeel zijn van het lesprogramma, zeker omdat ze voorzien in de overvloedige taalinput die ook voor andere vlakken van de tweedetaalverwerving belangrijk is. Bovendien leent extensief lezen zich goed voor thuis lezen, waartoe taalleerders mogelijk meer geneigd zijn als het leesmateriaal van goede kwaliteit is en past bij de leesbehoefte. Omdat lezen uiteindelijk een individuele zaak is, moet er ook individuele aandacht zijn voor het taalniveau, het leesdoel, de motivatie, de verwachting en de boekensmaak van een leerder. Daar is een begeleidende taak voor NT2-docenten weggelegd. Misschien kan dit onderzoek hen en NT2-lezers ondersteunen bij het maken van een passende boekenkeuze die helpt om de woordenschat te ontwikkelen en leesplezier te beleven.

LITERATUUR

Anthony, L. (2023). *AntWordProfiler* (version 2.1.0) [Computer software]. Waseda University.

Bolger, D., Balass, M., Landen, E., & Perfetti, C. A. (2008). Context variation and definitions in learning the meanings of words: an instance-based learning approach. *Discourse Processes*, 45(2), 122–159.

Bossers, B. (2022). Woordenschat. In F. Kuiken & S. Andringa (Red.), *Handboek Nederlands als tweede taal* (pp. 297–334). Coutinho.

Bruner, J. (1986). *Actual minds, possible worlds*. Harvard University Press.

Cohen, A., & Macaro, E. (2007). *Language learner strategies: 30 years of research and practice*. Oxford University Press.

College voor Toetsen en Examens. (z. d.) *Wat is het Staatsexamen NT2? Geraadpleegd op 22 september 2023*, van <https://www.staats-examensnt2.nl/over-het-staatsexamen-nt2/wat-is-het-staatsexamen-nt2>

Crossley, S., Louwerse, M., Mccarthy, P., & McNamara, D. (2007). A linguistic analysis of simplified and authentic texts. *Modern Language Journal*, 91(1), 15–30.

Crossley, S., Yang, H., & McNamara, D. (2014). What's so simple about simplified texts? A computational and psycholinguistic investigation of text comprehension and text processing. *Reading in a Foreign Language* 26(1), 92–113.

Day, R., & Bamford, J. (1998). *Extensive reading in the second language classroom*. Cambridge University Press.

Dolmans, R. (2015). *De uitgangspunten, opbouw en inhoud van de Woordenschat Toets Nederlands*. Boom.

Elgort, I., & Warren, P. (2014). L2 vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning*, 64(2), 365–414.

Elgort, I., Brysbaert, M., Stevens, M., & Assche, E. van (2018). Contextual word learning during reading in a second language: an eye-movement study. *Studies in Second Language Acquisition*, 40(2), 341–366.

Feng, Y., & Webb, S. (2020). Learning vocabulary through reading, listening, and viewing: which mode of input is most effective? *Studies in Second Language Acquisition*, 42(3), 499–523.

Gardner, D. (2013). *Exploring vocabulary: language in action*. Routledge.

González-Fernández, B., & Schmitt, N. (2017). Vocabulary acquisition. In S. Loewen & M. Sato (Eds.), *The Routledge Handbook of Instructed Second Language Acquisition* (pp.280–298). Routledge.

Goodman, K., & Freeman, D. (1993). What's simple in simplified language. In M. Tickoo (Ed.), *Simplification: theory and application* (pp. 69–76). Southeast Asian Ministers of Education Organization (Singapore). Regional Language Centre.

Grabe, W. (2009). *Reading in a second language: moving from theory to practice*. Cambridge University Press.

Grabe, W., & Stoller, F. (2011). *Teaching and researching reading*. Routledge.

Hiebert, E., & Kamil, M. (2005). *Teaching and learning vocabulary: bringing research to practice*. Lawrence Erlbaum Publishers.

Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 8(2), 689–696.

Honeyfield, J. (1977). Simplification. *TESOL Quarterly*, 11, 431–440.

Horst, M. (2005). Learning L2 vocabulary through extensive reading: a measurement study. *The Canadian Modern Language Review*, 61, 355–382.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: acquiring second language vocabulary through reading. *Reading in a Foreign Language*, 11, 207–223.

Hu, M., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13, 403–430.

Inclusion Europe. (2009). *Information for all. European standards for making information easy to read and understand* [Brochure]. Geraadpleegd op 10 januari 2024, van https://www.inclusion-europe.eu/wp-content/uploads/2017/06/EN_Information_for_all.pdf

Iperen, R. van (2022). 't Hooge Nest. Uitgeverij Eenvoudig Communiceren.

Javourey-Drevet L., Dufau, S., François, T., Gala, N., Ginestié, J., & Ziegler, J.-C. (2022). Simplification of literary and scientific texts to improve reading fluency and comprehension in beginning readers of French. *Applied Psycholinguistics*, 43(2), 485–512.

Land, J. (2009). *Zwakke lezers, sterke teksten? Effecten van tekst- en lezerskenmerken op het tekstbegrip en de tekstwaardering van umboleerlingen* (Dissertatie Universiteit Utrecht). Eburon.

Larsen-Freeman, D. (2002). *Techniques and principles in language teaching*. Oxford University Press.

Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: from humans to thinking machines* (pp.316–323). Multilingual Matters.

Laufer, B., & Ravenhorst-Kalovski, G. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22, 15–30.

Liu, N., & Nation, P. (1985). Factors affecting guessing in context. *RELC Journal*. 16(1), 33–42.

Long, M., & Ross, S. (1993). Modifications that preserve language and content. In M. Tickoo (Ed.), *Simplification: theory and application* (pp.29–52). Southeast Asian Ministers of Education Organization (Singapore). Regional Language Centre.

McLaughlin, B. (1987). *Theories of second language learning*. Edward Arnold.

Nagy, W., Herman, P., & Anderson, R. (1985). Learning words from context. *Reading Research Quarterly*, 20, 233–253.

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

Nation, P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Languages Review*, 63, 59–82.

- Nation, P. (2007). The four strands. *Innovation in Language Learning and Teaching*, 1(1), 1-12.
- Nation, P. (2009). *Teaching ESL/EFL reading and writing*. Routledge.
- Nation, P. (2013). Materials for teaching vocabulary. In B. Tomlinson (Ed.), *Developing materials for language teaching* (pp.351-364). Bloomsbury.
- Nation, P. (2014). Which words do you need? In J. Taylor (Ed.), *The Oxford Handbook of the Word* (pp.568-581). Oxford Handbooks.
- Nation, P., & Wang, M. (1999). Graded readers and vocabulary. *Reading in a Foreign Language*, 12, 355-380.
- Nation, P., & Waring, R. (2020). *Teaching extensive reading in another language*. Routledge.
- Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: an eye-tracking study. *Studies in Second Language Acquisition*, 38, 97-130.
- Phillips, M., & Shettlesworth, C. (1988). How to arm your students: a consideration of two approaches to providing materials for ESP. In J. Swales (Ed.), *Episodes in ESP: a source and reference book on the development of English for science and technology* (pp.105-111). Pergamon Press.
- Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: a case study. *Reading in a Foreign Language*, 18(1), 1-28.
- Raad van Europa (2022). *Gemeenschappelijk Europees Referentiekader voor talen: leren, onderwijzen, beoordelen – aangevulde uitgave*. Nederlandse Taalunie.
- Rets, I., & Rogaten, J. (2021). To simplify or not? Facilitating English L2 users' comprehension and processing of open educational resources in English using text simplification. *Journal of Computer Assisted Learning*, 37(3), 705-717.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge University Press.
- Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, 12, 329-363.
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, 95, 26-43.
- Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47(4), 484-503.
- Shook, D. (1997). Identifying and overcoming possible mismatches in the beginning reader-literary text interaction. *Hispania*, 80(2), 234-243.
- Silfhout, G. van (2014). *Fun to read or easy to understand? Establishing effective text features for educational texts on the basis of processing and comprehension research*. (Dissertatie Universiteit Utrecht). LOT.
- Sloot, K. van der, Hendrickx, I., Gompel, M. van, Bosch, A. van den, & Daelemans, W. (2018). *Frog, a natural language processing suite for Dutch* [Computer software]. Radboud University.
- Song, T., & Reynolds, B. (2022). The effect of lexical coverage on L2 learners' reading comprehension of narrative and expository genres. *Journal of English for Academic Purposes*, 59. 101154.
- Stichting Makkelijk Lezen. (z. d.) *Het ML-keurmerk voor makkelijk lezen boeken, uitgangspunten*. Geraadpleegd op 20 september 2023, van <https://www.stichtingmakkelijklezen.nl/uitgangspunten>
- Stokmans, M., & Wolters, R. (2019). *Lezen doe je samen*. Eburon.
- Subramaniam, S., & Zainal, Z. (2016). The effect of extensive reading on vocabulary acquisition. *LSP International Journal* 3(1), 63-82.
- Suk, N. (2017). The effects of extensive reading on reading comprehension, reading rate, and vocabulary acquisition. *Reading Research Quarterly*, 52(1), 73-89.
- Tiberius, C., & Schoonheim, T. (2014). *A frequency dictionary of Dutch: core vocabulary for learners*. Routledge.
- Tweissi, A. (1998). The effects of the amount and the type of simplification on foreign language reading comprehension. *Reading in a Foreign Language*, 11, 191-206.
- Waring, R., & Nation, P. (2004). Second language reading and incidental vocabulary learning. *Angles on the English-speaking World*, 4(4), 11-23.
- Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, 20, 232-245.
- Webb, S., & Chang, A. (2015). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition*, 37, 651-675.
- Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford University Press.
- Widdowson, H. (1978). *Teaching language as communication*. Oxford University Press.
- Yano, Y., Long, M., & Ross, S. (1994). Effects of simplified and elaborated texts on foreign language reading comprehension. *Language Learning*, 44, 189-219.

JET DOEDEL is hertaler van (literaire) romans en linguïst op het gebied van laaggeletterdheid en tweedetaalverwerving.
E-mail: jet@doedel.nl