

# Woordenschatgraven

## Over frequentielijsten en vocabulaireverwerving

MICHAEL BUIJKX

*Summary:* Many vocabulary lists in Dutch coursebooks are indirectly based on frequency research, and on the assumption that any particular selection of Latin literature might benefit from a representative list. However, the selection of merely 50 pages of Latin for Dutch senior high school education exams does not hold up to both linguistic and statistical logic. As a practical work-around a dichotomy is suggested between ‘content words’ and ‘structure words’. Whereas the latter (conjunctions, prepositions, pronouns and adverbs with a general meaning) can acceptably be selected on a frequency basis, the selection of the former (nouns, adjectives and verbs) can be tailored to the specific content of a text corpus with the help of the ‘Collatinus’ open source application. Statistical samples are taken from eight recent final examination programs and the criteria for a tailor-made selection are illustrated with Cicero’s *Pro Roscio* (national exams 2021). Various classroom activities and consequences for editing educational resources are presented.

### 1 Inleiding

Woorden leren in twee klassieke talen is een tijdrovende investering. Tegenover gymnasiasten wordt dit vooral gelegitimeerd met het rendement in de bovenbouw bij de taal van hun keuze: docenten en leerlingen vertrouwen erop dat de geleerde woorden voldoende dekkingpercentage zullen genereren. Dit uitgangspunt zou getoetst kunnen worden als al in de onderbouw met zekerheid vaststond welke teksten een bepaald cohort later precies gaat lezen.

In dit artikel wordt betoogd dat het leren van bestaande frequentielijsten te weinig zin heeft en het stelt als alternatief een gerichtere manier voor om leerwoorden te selecteren en aan te bieden. Allereerst wordt het fenomeen ‘basiswoordenlijst’ tegen het licht gehouden. Statistische en linguïstische wetmatigheden oefenen namelijk grote druk uit op de bruikbaarheid van een frequentie criterium wanneer het totale bovenbouwcorpus niet veel groter is dan 10.000 woorden (20 pagina’s Oxfordtekst voor het Centraal Examen [CE] en 30 pagina’s voor het schoolexamen [SE]). Omdat de spreiding en frequentie van de verschillende woordsoorten onderling sterk verschilt, wordt

voorgesteld een praktische tweedeling aan te brengen tussen *inhoudswoorden* (substantiva, adjectiva en verba), die in hoge mate afhangen van de inhoud van een tekstpensum, en *structuurwoorden* (alle overige woordsoorten).

De opbrengst van zo'n tweedeling wordt geïllustreerd door het dekkingspercentage in acht recente examenpensa Latijn te analyseren met behulp van de *open source* applicatie Collatinus. Door de te leren woorden als maatwerk te selecteren en in enkele vaste en flexibele categorieën te ordenen kan het memoriseren en consolideren van de woordenschat efficiënter worden ingezet om de lectuur te ondersteunen. Deze aanpak wordt met het pensum voor het CE 2021 uit de *Pro Roscio* van Cicero geïllustreerd.

In het tweede deel van dit artikel worden diverse voorstellen gedaan voor de lespraktijk omtrent woordenschatverwerving. Eerst moet echter duidelijk worden op welke aannames frequentiewoordenlijsten gebaseerd zijn en welke wetmatigheden de woordenschat in een tekstcorpus beïnvloeden.

## 2 Taal- en rekenkunde bij frequentieonderzoek

### 2.1 Veelbelovend computeronderzoek

In 1974 publiceerde het *Laboratoire d'Analyse Statistique des Langues Anciennes* (LASLA) in Luik een veelgebruikte frequentieanalyse van Latijnse woorden in een representatief geacht corpus.<sup>1</sup> In totaal werden 794.662 woorden toegeschreven aan 13.077 *lemmata*.<sup>2</sup> Het vertrouwen in de mogelijkheden van de computer om betrouwbare gegevens te genereren was groot:

Nous croyons que l'ampleur de la documentation que nous possédons nous permet de présenter un dictionnaire fréquentiel de la langue latine hautement représentatif de l'usage courant.<sup>3</sup>

Analyses zoals die van LASLA werden gebruikt bij het uitgeven van zogenaamde basiswoordenlijsten als een soort hitparade van 'de' meest frequente woorden.<sup>4</sup> Deze basiswoordenlijsten werden vervolgens doelbewust verwerkt in teksten van onderbouwleergangen. Onder invloed van deze schoolboeken

1 <http://web.philo.ulg.ac.be/lasla/>.

2 <https://www.stilus.nl/woorden/lasla1974.pdf> en <https://www.stilus.nl/woorden/lasla1981.pdf>. In dit artikel wordt met *woorden* bedoeld: alle woorden (*tokens*) zoals ze in verschillende vormen voorkomen in een tekst; met *lemma* wordt de woordenboekvorm bedoeld (*type*) waarop de verschillende vormen en voorkomens van een woord kunnen worden herleid.

3 Zie Delatte *et al.* (1981: I).

4 De werkwijze bij LASLA is uiteraard veel professioneler dan hier gesuggereerd wordt; in dit artikel gaat het om oude lijsten die ingeburgerd zijn in basiswoordenlijsten en onderbouwleergangen. Bij LASLA kan een eigen inlogcode worden aangevraagd waarmee bijvoorbeeld de frequentieanalyse van de hele *Pro Roscio* kan worden gedownload: <http://web.philo.ulg.ac.be/lasla/opera-latina/>.

leek er een consensus te zijn ontstaan welke woorden wel of niet geleerd moeten worden, ze werden ‘basiswoorden’ genoemd. Afgezien van blokken ‘basiswoorden’ per tekst verschenen er ook ‘basiswoordenlijsten’ geordend op alfabet, frequentie, thema, verbuigingsgroep etcetera.<sup>5</sup>

De *Basiswoordenlijst Latijn* (Babeliowsky *et al.* 1977) bestaat uit 1765 woorden met daarnaast aparte micro-lexica per auteur. Veel woorden zijn voorzien van een authentiek citaat dat is geattesteerd in de tien Latijnse auteurs die in de toenmalige schoolpraktijk het meest aan bod kwamen.<sup>6</sup> Het verwachte rendement werd als volgt beredeneerd:

Daarbij is als norm aangehouden dat gemiddeld op elke 100 woorden die bij deze auteurs worden aangetroffen er ten minste 85 in de basiswoordenlijst moeten voorkomen.<sup>7</sup>

Het recente *Essential Latin Vocabulary* bestaat uit de ‘1425 most common words occurring in the actual writing of over 200 Latin authors’, die zowel systematisch als thematisch als alfabetisch geordend zijn, met een frequentieberekening per woord tot op twee decimalen nauwkeurig. Williams belooft zelfs: ‘The text presents 1425 words that allow a student to comprehend about 95% of all the vocabulary they will ever see in an actual Latin text’.<sup>8</sup> Dit klinkt te mooi om waar te zijn, en dat is het ook.

In geen van de basiswoordenlijsten of onderbouwleergangen klinkt enige twijfel door of het uitgangspunt van frequentieonderzoek wel voldoende representatief kan zijn om te beslissen welke woorden een leerling moet memoriseren. Nochtans waren er al lang voor de digitale revolutie inzichten beschikbaar uit de taalkunde en statistiek die hier principiële vragen bij stellen.

## 2.2 *Woordfrequentie in teksten*

Het leerplan Latijn van de VCN uit 1982 trok in de paragraaf over woordenschat conclusies op basis van de LASLA-analyse van Caesars *De bello Gallico*: dit werk bevat 45.500 woorden, die herleid kunnen worden op slechts 2800 verschillende *lemmata*. Caesars woordenschat is relatief klein en (dus) efficiënt. De 100 meest frequente unieke woorden dekken bijna de helft (22.000 woorden) af; de 500 meest frequente 75%; de 750 meest frequente 80%.<sup>9</sup> Het leerplan Latijn stelde dat van die 2800 unieke woorden ‘de meeste uiteraard

5 Bijvoorbeeld Hupperts en De Jong (1998); Wellink (1997); Brouwer *et al.* (1991); Kanters (1979); Lenaers (1997).

6 Caesar, Cicero (redevoeringen en filosofische werken), Sallustius, Livius, Seneca (brieven), Tacitus, Catullus, Vergilius, Horatius (oden) en Ovidius.

7 Zie Babeliowsky *et al.* (1975: 1).

8 Zie Williams (2012: x).

9 Zie Van Amerongen *et al.* (1986: 16).

meer dan eenmaal voorkomen.<sup>10</sup> Op zich is dit correct, want de optelsom van alle woorden die twee keer of vaker voorkomen overtreft het aantal woorden dat slechts één keer voorkomt. Maar het echte probleem wordt juist gevormd door de woorden die slechts één keer voorkomen. Dat zijn namelijk niet per se ongebruikelijke woorden, ook van zogenaamde ‘basiswoorden’ komen er in ieder concreet tekstpensum vele slechts één keer voor.

George Kingsley Zipf had al in 1935 een wetmatigheid aangetroffen in de frequentie van woorden in teksten. Het meest voorkomende woord komt twee keer zo vaak voor als het op één na meest voorkomende woord en drie keer zo vaak als het op drie na meest voorkomende woord tot aan het minst gebruikte woord in de tekst.<sup>11</sup> Zodoende is er in elk tekstcorpus een zeer grote hoeveelheid woorden die slechts één keer voorkomt terwijl een vrij kleine groep van hoogfrequente woorden een buitenproportioneel groot percentage vormt.

De totale omvang van een tekstcorpus bepaalt het aantal posities en dus de speelruimte voor herhaling en spreiding van woorden. William Bull kende de wet van Zipf nog niet toen hij in 1949 een zeer leesbaar artikel schreef over de fundamentele logische, linguïstische en statistische problemen wanneer leerwoordenlijsten op frequentieonderzoek worden gebaseerd. Door vergelijking van Spaanse, Engelse en Latijnse woordfrequenties kwam hij tot een *natural frequency*: de wisselwerking tussen culturele en linguïstische factoren die bepalend is voor de frequentie van woorden in een tekst. Diederich had in een corpus van 3 omvangrijke bloemlezingen al geconstateerd dat slechts 21 Latijnse lemmata 25% van elke Latijnse tekst vormen.<sup>12</sup> Op een pensum van 10.000 woorden levert het leren van slechts 21 woorden dus circa 2500 keer de vreugde der herkenning op.<sup>13</sup> Bull belicht de keerzijde van dit goede nieuws: er zijn dan nog maar 7500 posities over voor alle andere woorden. Dus als de 100 meest frequente woorden 50% van een tekst afdekken, dan zijn er nog maar ten hoogste 5000 posities over voor alle andere dan die 100 woorden, enzovoorts. Zo ontstaat er een grote spreiding van de andere dan de meest frequente woorden, die vanwege de wet van Zipf in steeds kleinere aantallen optreden. Het feit dat een gering aantal woorden (*esse*, de pronomina, preposities en conjuncties) om taalkundige redenen hoogfrequent is, verkleint de beschikbare ruimte voor de woorden die het onderwerp van de tekst dragen.

10 Volgens de Perseus *Vocabulary Study Tool* komen in *De Bello Gallico* slechts 638 van de 4535 unieke woorden éénmaal voor en bedraagt het totaal aantal woorden 51.295.

11 Het raadsel van de wet van Zipf is recent verklaard met behulp van de wisselwerking tussen semantiek en syntaxis, zie Lestrade (2017) en Jansen en Jaspers (2017). Als formule luidt de wet: de rang van een woord  $r$  maal de frequentie van een woord  $f$  is constant.

12 Zie Diederich (1939): *-que, qui/quis, et, sum, in, is, hic, non, cum, ad, ille, ut, omnis, ab, suus, dico, nec, de, sed, ipse, si*.

13 In dit artikel wordt voorbijgegaan aan de vele verschillende betekenissen van die hoogfrequente woorden. Voor Grieks trekt Major (2008) vergelijkbare conclusies over de kern van de woordenschat.

Binnen die beperkte ruimte is het zo steeds onwaarschijnlijker dat woorden vaker dan één of twee keer voorkomen. Hoe kleiner de omvang van het tekstcorpus is, hoe dwingender dit effect optreedt.<sup>14</sup> Diederich vond 1471 *essential words* die gemiddeld 83,5% van de teksten afdekken.<sup>15</sup> Door velen werd gedacht dat dit dekkingspercentage geëxtrapoleerd kon worden ongeacht de omvang van het corpus. Ten onrechte: in een corpus van 10.000 woorden zou er dan theoretisch nog maar ruimte zijn voor 1650 andere woorden; in een corpus van 2000 woorden kunnen onmogelijk alle 1471 *essential words* voorkomen. Het is dus mede van de omvang van het uiteindelijk te lezen pensum afhankelijk of het door een basiswoordenlijst beloofde dekkingspercentage geattesteerd kan worden.

### 2.3 *Frequentiepatronen in zinnen*

Naast de beperkingen die samenhangen met de omvang van een tekstcorpus is een tweede probleem inherent aan het feit dat de woorden in een tekst nooit ‘los’ voorkomen maar binnen de structuur van zinnen. Dit heeft grote gevolgen voor de frequentie van de verschillende woordsoorten. Elke woordsoort volgt zijn eigen linguïstische logica in een concrete zin. In vrijwel alle zinnen is er ten minste één substantivum nodig (vervangbaar door een pronomen) en een verbum. Deze woordsoorten zijn dus, vanwege de zinsbouw, in elke tekst oververtegenwoordigd, waarbij substantiva het leeuwendeel vormen. Acties die gedaan of ondergaan worden kunnen met meerdere substantiva verbonden worden en dus komen verba in aantal op de tweede plaats. Eigenschappen van die substantiva komen op de derde plaats, ook weer omdat er meer combinaties zijn waarin een adjectivum een substantivum kwalificeert dan omgekeerd.<sup>16</sup> Een aantal adverbia kan in een zin op verschillende niveaus modifieren (vergelijk *multo currit* en *multo celerius currit*) en dus zijn er in totaal minder adverbia nodig dan de woordsoorten die zij in een zin modifieren.<sup>17</sup>

### 2.4 *Woordsoorten en verdringing*

Binnen het taalsysteem als geheel is er ook een inhoudelijke verklaring voor de sterk uiteenlopende aantallen van de verschillende woordsoorten. Sommige woorden worden vaak gebruikt omdat er weinig alternatieven beschikbaar zijn, terwijl binnen andere woordsoorten er zoveel onderlinge concurreren.

14 Ten tijde van de *Basiswoordenlijst* van Babeliowsky *et al.* (1979) werden er 100 pagina's Oxfordtekst voorgeschreven voor het SE, door het leerplan Latijn (Van Amerongen *et al.* 1986) 70 pagina's, sinds de invoering van de tweede fase in 1998 50 pagina's.

15 Zie Williams (2012: x).

16 Zie Bull (1949: 476). De onderlinge verhouding tussen het aantal adjectiva en verba is wel afhankelijk van het indelingscriterium: deelwoorden als *amans* of *amatus* kunnen de balans beïnvloeden.

17 Zie Bull (1949: 475-6).

rentie tussen de woorden is dat een hoge frequentie van één concreet woord zeldzaam is.

Van alle woordsoorten is bij de substantiva de behoefte aan verschillende woorden het grootst, niet eens zozeer door synoniemen met semantische verschillen (*timor, metus, pavor*) maar om de werkelijkheid zo precies mogelijk te kunnen beschrijven (*gladiator, retiarius, murmillo, bestiarius*). Ook om stilistische redenen en bij poëzie vanwege metrum en klank is er soms een groot repertoire aan woorden voor één begrip nodig (*lymphā, liquor, humor, unda, latex*). Congruerende woorden vervullen beschrijvende of preciserende functies en variëren van woorden voor hoeveelheid (*multi, tantus*) of bezit (*meus*) tot veel specifiekere adjectiva die sterk inhoudsafankelijk zijn (*argenteus*).<sup>18</sup> Voor een ontkenning zijn er niet veel alternatieven nodig (*non, haud, neque*). Ook van de preposities en conjuncties zijn er per begrip (*hoewel, naar*) slechts enkele woorden nodig die dus elk veelvuldig voorkomen.<sup>19</sup>

Voor de frequentie van woordsoorten in een tekst betekent dit het volgende. Er zijn zoveel verschillende substantiva dat daarvan de kans dat één bepaald woord vaak gebruikt wordt het kleinst is. De taalgebruiker heeft hier de meeste alternatieven voorhanden en wil de werkelijkheid zo precies mogelijk beschrijven. Daarentegen is het aantal verschillende preposities beperkt, waardoor bijvoorbeeld het voorzetsel *in* vaak voorkomt in elke tekst. De voorkeur van een schrijver heeft hier nauwelijks invloed op: het is nagenoeg onmogelijk om een tekst zonder preposities, conjuncties of pronomina te schrijven. Ook *esse* (koppelwerkwoord en hulpwerkwoord bij de passieve perfectā) legt beslag op het totaal aantal woorden in elke langere Latijnse tekst.

Dit soort taalkundige wetmatigheden hebben nog een ander gevolg voor frequentielijsten waarin de woorden van tekstcorpora worden gerangschikt. Onder de meest frequente 25 woorden komt geen enkel substantivum voor. Maar wordt de groep van de eerste 1000 woorden bekeken, dan domineren de 357 substantiva daarin al zo sterk dat er voor niet meer dan 133 van de meest voorkomende adjectiva plaats overblijft.<sup>20</sup> In elk volgende duizendtal geldt dit verdringingseffect opnieuw door het grote aantal substantiva dat in de zinnen van een tekstcorpus nodig is. Daardoor kan in de eerste 1000 woorden van de LASLA-lijst een woord als *foramen* op de 972<sup>ste</sup> plaats voorkomen terwijl *albus* erbuiten valt.<sup>21</sup> De onderlinge volgorde van de woordsoorten in een frequentielijst wordt dus vervormd omdat de meeste zinnen nou eenmaal niet zonder een verbum en een substantivum kunnen. Het werkt daarom beter om

18 Hulstaert (2016: 20) maakt een zinnig onderscheid tussen beschrijvende adjectiva en onderscheidende adjectiva zoals stofnamen (*aureus*), volksnamen (*Romanus*) en persoonsnamen (*paternus*).

19 Zie Bull (1949: 475).

20 Zie Bull (1949: 483). Aantallen op basis van de LASLA top 1000.

21 Deze eer dankt *foramen* overigens niet aan Pyramus en Thisbe of medisch Latijn maar aan Lucretius.

elke woordsoort op de eigen merites in een pensum te beoordelen dan om erens, bijvoorbeeld na de eerste 1000 woorden, een caesuur te leggen.

### 2.5 *Semantische frequentiepatronen*

Naast linguïstische noodzaak verklaren ook de keuzes van taalgebruikers en tekstschrijvers waarom sommige woorden over- of juist ondervertegenwoordigd zijn. Bull onderscheidt op basis van hun gebruikssituatie negen categorieën woorden in een taal die steeds andere frequentiepatronen opleveren. Naast de al besproken *inescapable words* (*in, esse, et*) verklaren zijn *rare-common* en *common-rare words* het conceptuele probleem het best.<sup>22</sup> *Common-rare* noemt Bull de woorden uit het dagelijks leven (*keuken* en *plassen*) die elk kind al kent en vaak gebruikt – maar die niet per se hoogfrequent zijn in geschreven teksten. *Rare-common* noemt hij de vele woorden als *ontstoppen* of *lancet* die elke loodgieter of chirurg voortdurend nodig heeft maar die buiten dat specialisme nauwelijks gebruikt worden. Zijn conclusie is dat er enerzijds hoogfrequente woorden zijn die uit linguïstische noodzaak in elke tekst vaak voor moeten komen, buiten de wil van de spreker om, maar die op zich nauwelijks bijdragen om de strekking van een tekst te begrijpen. Anderzijds wordt begrip van de inhoud van een tekst grotendeels gedragen door substantiva en verba. Maar omdat er zoveel onderlinge concurrentie is kunnen de meeste van deze woorden niet hoogfrequent zijn. Het grootste probleem is dat woordfrequentie dermate afhangt van het onderwerp van de tekst dat er gewoonlijk weinig overlap is tussen de inhoudswoorden die we vinden in verschillende teksten.

### 2.6 *Dekkingsgraad en tekstbegrip*

De hierboven besproken linguïstische factoren hebben ook consequenties voor de effectiviteit van de gegroeide praktijk op scholen om ‘basiswoorden’ te laten leren. Het tellen van woorden in andere teksten dan die een leerling feitelijk zal lezen leidt tot een investering in leerwerk die onmogelijk rendement kan opleveren bij de lectuur. Er staan te veel taal- en rekenkundige wetten in de weg.

Bovendien is het effect van een basiswoordenschat op de lectuur van teksten beperkter dan voorheen werd aangenomen. In de twintigste eeuw ging men ervan uit dat een dekkingsgraad van 85% voldoende zou zijn om de onbekende woorden in een tekst af te kunnen leiden uit de context. Dit percentage zou voor Latijn bereikt kunnen worden als circa 1800 woorden gekend werden en als er aangeleerd was hoe andere woorden hiervan afgeleid worden,

22 Verder onderscheidt Bull (1949: 479): *linguistic imperatives* (woorden van ruimte, tijd, aantal, volgorde, enzovoort); *socio-psychological terminology* (familierelaties, begroetingen, emoties, en dergelijke); *rare-rare words*; *archaic words*; *personal inventions* en *taboos*.

zoals samengestelde werkwoorden en wortels met prefixen en suffixen.<sup>23</sup> Als er aandacht besteed wordt aan principes van woordvorming zou het beoogde dekkingspercentage bereikt kunnen worden met een overzichtelijk aantal stammen, die elk immers in grote ‘woordfamilies’ herkenbaar zijn (bijvoorbeeld/regel => *regēre, rex, regina, rectus*, enzovoort).

In de 21<sup>ste</sup> eeuw is het aantal onderzoeken naar zowel woordenschatverwerking als naar tekstbegrip exponentieel toegenomen.<sup>24</sup> Tegenwoordig gaat men ervan uit dat ten minste 95% van de woorden in een tekst bekend moet zijn om bij eerste lezing tot tekstbegrip te komen, en zelfs 98% om met plezier te kunnen lezen.<sup>25</sup> Dit onderzoek betreft echter voornamelijk Engelse teksten. In flecterende talen kunnen de geleerde woorden ook nog eens in allerlei uiteenlopende woordvormen verschijnen. Zelfs als leerlingen alle woorden in hun onderbouwleergang, deels met meer dan één betekenis en extra informatie (zoals genitivusvorm, perfectumstam, informatie over de constructie) geleerd en regelmatig herhaald hebben, is herkenning in teksten nog niet gegarandeerd.

### 2.7 *Analyse van de dekkingsgraad van een frequentielijst in acht examenpensa*

Hoe dekkingsgraad functioneert in het gymnasiumonderwijs kan worden gedemonstreerd aan acht recente examenpensa Latijn (ieder 20 pagina’s Oxfordtekst). In onderstaande analyse heb ik de 1000 meest frequente woorden volgens de LASLA-lijt gebruikt. Deze 1000 woorden van LASLA zijn als Excelfile gepubliceerd op [stilus.nl](http://stilus.nl) op volgorde van frequentie, en daar van Nederlandse betekenissen, aanvullingen en woordsoorten voorzien.<sup>26</sup>

Voor het uitvoeren van de analyse gebruikte ik de *open source* applicatie Collatinus.<sup>27</sup> Collatinus heeft verschillende functionaliteiten en kan woordfrequentie in een concrete digitale Latijnse tekst zowel analyseren als visualiseren.<sup>28</sup>

De analyse van de dekkingsgraad van een woordenlijst in een concrete

23 Lenaers (1997: 174) geeft een lijst van 33 wortels die ‘een manier om te zijn’ aangeven. Zij hebben alle een werkwoord op -ēre (*candēre*); een inchoativum op -escēre (*candescēre*); een adjectivum op -idus (*candidus*) en een substantivum op -or (*candor*). Vergelijk *calidus, horridus, humidus, pavidus, rigidus, validus, stupidus, timidus*, enzovoort.

24 Nation (2013) is een overzichtsstudie over onderzoek naar woordenschat.

25 Zie Webb en Nation (2017: 12); Hu en Nation (2000); Nation (2013: 352).

26 Naar de eerste 1000 woorden uit de LASLA-lijt op [stilus.nl](http://stilus.nl) wordt in dit artikel verder gerefereerd als: LASLA top 1000, zie Nellissen (2018). Veel dank is verschuldigd aan Leo Nellissen voor al zijn onschatbare werk via [stilus.nl](http://stilus.nl).

27 Collega Jan Bart komt de eer toe dat hij Collatinus voor Nederlandse classici heeft ontsloten en de visualisatie door ‘verkleuring’ bij *biblissima* voor elkaar heeft gekregen, zie <https://outils.biblissima.fr/en/collatinus/>.

28 Zie paragraaf 3.3 voor de visualisatie van woordfrequentie, die tool is vooral handig in de lespraktijk.



tekst gaat met behulp van Collatinus als volgt. In Collatinus kan een kale lijst Latijnse woorden worden ingelezen die dan vergeleken wordt met de woorden in een concreet tekstpensum.<sup>29</sup> De 1000 woorden van de frequentielijst van LASLA zijn ingelezen en vergeleken met de vier meest recente examenpensa proza (2021, 2018, 2017, 2015); en vier pensa poëzie (2020, 2019, 2016, 2011), (zie tabel 1). Collatinus berekent automatisch het dekkingspercentage: de herkende woorden uit de ingelezen lijst gedeeld door het totaal aantal woorden.

Het leren van de LASLA top 1000 levert herkenning op van ten minste 2589 woorden (Vergilius 2016) en ten hoogste 3841 (*Pro Roscio* 2021). Het door Collatinus berekende dekkingspercentage komt tot drie maal toe uit op 67% voor Vergilius, op 70% voor Ovidius en schommelt rond de 75% voor proza.

Het beeld van deze dekkingspercentages wordt echter niet in gelijke mate veroorzaakt door de verschillende woordsoorten, waarvan een groot deel immers wordt gedekt door Diederichs 21 *inescapable words*. Binnen alle onderzochte pensa valt de woordenschat in twee groepen uiteen met eigen frequentiepatronen.

## 2.8 Tweedeling in woordsoorten

In het Latijn is het verschil tussen verbuigbare en onverbuigbare woorden goed bruikbaar om tot een tweedeling in woordsoorten te komen die zich qua frequentie verschillend gedragen.

Verbuig- en vervoegbaar zijn verba, substantiva, adjectiva en pronomina. De pronomina verwijzen naar een eerdergenoemd woord of begrip in de tekst en zijn daarom hoofdfrequent, maar ze hebben per definitie geen ‘eigen’ betekenis of semantische lading. Hoogfrequent is bijvoorbeeld ook het begrip /mens/ (de verzameling van alle substantiva die mensen aanduiden) of /doen/ (de verzameling van de meeste transitieve werkwoorden). Daarentegen is één concreet substantivum voor een mens, *praetor* of *uxor*, alleen dan hoofdfrequent in een pensum als de inhoud daar aanleiding toe geeft.

Van de onverbuigbare woorden zijn de meest voorkomende woorden (*ut*, *an*, *inter*) doorgaans korter dan minder frequent gebruikte (*profecto*, *improbabiliter*). Ze zijn ook abstracter van betekenis, en in de ervaring van leerlingen lastiger om te leren. Terwijl *rosa*, *bibere* of *splendidus* een visueel beeld oproepen, geldt dat voor *dum*, *ac* en *sine* niet. Ook bij de adverbia is er een glijdende schaal: de van adjectiva afgeleide adverbia als *feliciter* of *celerrime* hebben een overwegend semantische lading, terwijl frequente adverbia als *tunc* en *ibi* in veel gevallen (net als pronomina) verwijzen naar inhoudelijke

29 Hoe een set Latijnse woorden slim ontdaan wordt van macrons legt onze leerling Daan Boven uit in ‘Latijn: macrons weghalen in MS Word’, <https://youtu.be/l6dYAAA4jGQ>.

Tabel 1 Dekkingspercentage in acht examenpensa op basis van de LASLA top 1000

CE	pensum	Totaal aantal woorden (x)	Woorden uit de LASLA top 1000 (y)	Door Collatinus berekend dekkingspercentage (y/x)
2021	Cicero <i>Pro Roscio</i>	4943	3841	78%
2018	Cicero/Seneca	4571	3524	77%
2017	Livius	5030	3536	70%
2015	Cicero/Plinius	4906	3755	76%
2020	Vergilius	3938	2640	67%
2019	Ovidius	3897	2832	72%
2016	Vergilius	3838	2589	67%
2011	Vergilius	3941	2653	67%

informatie elders in de tekst.<sup>30</sup>

Om praktische redenen is ervoor gekozen de Latijnse woorden die in een pensum worden aangetroffen op basis van hun woordsoort te verdelen in twee groepen: inhoudswoorden en structuurwoorden. Inhoudswoorden zijn hier alle substantiva, adjectiva en verba, dus de verbuigbare woordsoorten zonder de pronomina. Alle overige woordsoorten, dus de onverbuigbare en de pronomina, worden hier onder de verzamelnaam structuurwoorden geschaard.<sup>31</sup> Beide groepen vereisen een eigen aanpak bij het leerwerk omdat ze zich anders verhouden tot de feitelijke inhoud van teksten en andere frequentiepatronen vertonen.

## 2.9 Inhoudswoorden

Bij de verba, substantiva en adjectiva is er sprake van morfologische variatie, semantische spreiding over een groot aantal lemmata en daardoor zeer grote statistische spreiding in teksten: omdat er zoveel van zijn, zijn er weinig inhoudswoorden die in een tekstpensum een hoge frequentie bereiken.<sup>32</sup>

30 In de taalkunde is het gebruikelijk alle adverbia bij de inhoudswoorden onder te brengen. Voor de langere adverbia is dit op zich beter (vergelijk bijvoorbeeld de Engelse *adverbs* op -ly) maar het primaire doel van de hier gehanteerde tweedeling is slechts het scheiden van hoogfrequente woorden met geringe semantische lading van de inhoudswoorden. De adverbia die van adjectiva zijn afgeleid verstoren dit beeld niet.

31 Wanneer woorden tot verschillende woordsoorten kunnen behoren, zoals *hic* of *satis*, komen ze doorgaans toch in de juiste helft terecht: *quod* en *cum* zijn in elk geval 'structuurwoord'; *multum* en *amicus* inhoudswoord. Relatief weinig ambigue woorden (*verum*) onttrekken zich aan de dichotomie.

32 Woorden als *fero*, *ago*, *res*, zijn hoogfrequent maar hebben wel veel verschillende betekenissen. Het leren van 'fero 1. dragen 2. brengen' helpt niet voor *in nova fert animus*, dan moet er gekozen worden uit 29 betekenisnummers in het woordenboek van Pinkster.

Wanneer de steekproef zo gering is als in de 50 pagina's OCT (10.000 woorden) voor de hele bovenbouw, kunnen de meeste inhoudswoorden niet vaker dan één of twee keer voorkomen. De kans dat een woord daadwerkelijk voorkomt in een concreet pensum hangt af van twee factoren: de 'algemeenheid' van de semantische lading van het woord en de feitelijke inhoud van de teksten. Algemene woorden zoals *agere, dicere, facere, res, vir, locus, novus, multi* en *magnus* zijn hoogfrequent ongeacht tekstsoort of inhoud. Ze worden in tekstverband verbonden met namen of concrete begrippen en kunnen daar veel verschillende verbindingen mee aangaan.<sup>33</sup> *Bibere* kan echter in de Romeinse cultuur met weinig andere begrippen verbonden worden dan *aquam, vinum* of omschrijvingen daarvan. De frequentie van deze woorden hangt vooral samen met de feitelijke inhoud van een tekst, en veel minder met genre, stijl, auteur of periode. Epische dichters gebruiken graag *aequor, altum* of het Griekse leenwoord *pelagus* en geschiedschrijvers liever *mare*. Maar als er in een bepaald pensum alleen op het land gevochten wordt of als Aeneas afdaalt in de onderwereld komt het hele begrip /zee/ niet voor. Alleen bij synoniemen zijn keuzes van de schrijver en stilistische of genreconventies relevant. Bijgevolg zijn microlexica van woorden die in het hele oeuvre van een bepaalde auteur vaak voorkomen weliswaar nuttig voor classici, maar voor leerlingen die na 50 pagina's Oxfordtekst hun eindexamen halen nauwelijks.<sup>34</sup>

Het is daarentegen wel zinvol om alle inhoudswoorden die drie keer of vaker voorkomen *binnen* het CE pensum te leren en/of consolideren. Ook voor het SE pensum is '≥3 keer' een werkbaar criterium als een docent zelf de teksten per klas of leerjaar op frequentie wil analyseren.<sup>35</sup>

### 2.10 Structuurwoorden

Alle andere woordsoorten dan substantiva, adjectiva en verba worden samen genomen in één restgroep van 'structuurwoorden'.<sup>36</sup> Zoals de 21 *inescapable words* al lieten zien, kan een Latijnse literaire tekst het moeilijk stellen zonder nevenschikkende voegwoorden of voorzetsels. Ook adverbia als *postea* en *illic* dienen, net als pronomina, om herhaling van steeds dezelfde namen of bijwoordelijke bepalingen te vermijden. Prozaschrijvers gebruiken meer

33 Frequente woorden als *res, gerere* of *maior* hebben 'los' een weinig specifieke semantische lading. In zinnen worden zij gebruikt in combinatie als *res publica, bellum gerere* of *mos maiorum* die wel een specifieke betekenis hebben. Dit artikel beperkt zich tot losse woorden en laat zogenaamde collocaties buiten beschouwing.

34 Zie Babeliowsky *et al.* (1975).

35 Ook een uitgever zou de verzamelde teksten van zijn onderbouwleergang met dit criterium kunnen doorrekenen.

36 Met opzet is niet gekozen voor *content words* en *function words*, zie Nation (2013: 308-9). In dit artikel wordt niet zozeer taalkundige precisie beoogd, maar een werkbaar criterium om vaste van flexibele lijsten te scheiden.

Tabel 2 Dekkingspercentage in acht examenpensa op basis van de 188 structuurwoorden in de LASLA top 1000

CE	pensum	Totaal aantal woorden (x)	188 structuurwoorden (y)	Door Collatinus berekend dekkingspercentage (y/x)
2021	Cicero <i>Pro Roscio</i>	4943	1978	40%
2018	Cicero/Seneca	4571	1797	39%
2017	Livius	5030	1511	30%
2015	Cicero/Plinius	4906	1953	39%
2020	Vergilius	3938	949	24%
2019	Ovidius	3897	1085	27%
2016	Vergilius	3838	897	23%
2011	Vergilius	3941	1063	26%

Cicero *Pro Roscio* c. 17-18, alleen de herkende structuurwoorden:

...ei .....cum duobus ...quorum ...in ...tria ...quas ...si tam ...quam ...Neque enim, ...Nam duo isti ..., quorum ...iste, qui ...eius modi: ...ac ...hic autem ...se ad eum ...quique ante hanc ...quod ... ipsum ...Nam cum hic ... ..., ...autem iste ...cum hic ...in ...cumque se ...iste autem ...ad ... ..a ...

Cicero *Pro Roscio* c.17-18, alleen de inhoudswoorden + namen:

Erant ...veteres inimicitiae ...Rosciis Amerinis, .....alterum sedere ...accusatorum subselliis video, alterum ... praedia possidere audio; ...inimicitias ...cavere potuisset, ...metuere solebat, viveret. ...iudices, iniuria metuebat. ...sunt T. Roscii, ...alteri Capitoni cognomen est, ...adest, Magnus vocatur, homines ...Alter ...palmarum vetus ...nobilis gladiator habetur, ...lanistam contulit, ...pugnam tiro esset ...sciam, ...magistrum scelere audaciaque superavit. ...Sex. Roscius esset Ameriae, T. ...Roscius Romae, ...filius adsiduus ...praediis esset ...voluntate patris rei familiari vitaeque rusticae dedisset, ...frequens Romae esset, occiditur ...balneas Pallacinas rediens cena Sex. Roscius.

Cicero *Pro Roscio* c. 17-18, alle 127 woorden:

Erant ei veteres inimicitiae cum duobus Rosciis Amerinis, quorum alterum sedere in accusatorum subselliis video, alterum tria huiusce praedia possidere audio; quas inimicitias si tam cavere potuisset, quam metuere solebat, viveret. Neque enim, iudices, iniuria metuebat. Nam duo isti sunt T. Roscii, quorum alteri Capitoni cognomen est, iste, qui adest, Magnus vocatur, homines eius modi: Alter plurimarum palmarum vetus ac nobilis gladiator habetur, hic autem nuper se ad eum lanistam contulit, quique ante hanc pugnam tiro esset quod sciam, facile ipsum magistrum scelere audaciaque superavit. Nam cum hic Sex. Roscius esset Ameriae, T. autem iste Roscius Romae, cum hic filius adsiduus in praediis esset cumque se voluntate patris rei familiari vitaeque rusticae dedisset, iste autem frequens Romae esset, occiditur ad balneas Pallacinas rediens a cena Sex. Roscius.

*Afb. 1 Caput 17-18 uit Cicero Pro Roscio.*

structuurwoorden dan dichters, die juist veel verschillende inhoudswoorden (bijvoorbeeld synoniemen en poëtische omschrijvingen) nodig hebben.

In de LASLA top 1000 blijven, na aftrek van de substantiva, adjectiva en verba, 188 structuurwoorden over. De dekkingsgraad van deze 188 structuurwoorden in de examenpensa levert het volgende beeld op (zie tabel 2): een kwart van de woorden bij de vier pensa poëzie en 30 à 40 % van de pensa proza worden afgedekt door de 188 structuurwoorden.

Hoewel sommige structuurwoorden vaker voorkomen dan andere, staat hun frequentie in teksten relatief los van de feitelijke inhoud. Afbeelding 1 laat zien dat als in een passage (Cicero, *Pro Roscio* 17-18) alle inhoudswoorden worden weggelaten er geen enkel tekstbegrip mogelijk is, maar omgekeerd kan een globaal idee van de inhoud wel gevormd worden zonder de structuurwoorden.

Op grond van deze gegevens is de conclusie gerechtvaardigd dat het selecteren van woorden om te leren door uitgevers en docenten efficiënter kan worden georganiseerd. Door een vaste lijst van de 188 meest voorkomende structuurwoorden te combineren met een flexibele lijst van die inhoudswoorden die binnen een concreet pensum drie keer of vaker voorkomen, kan een optimale verhouding bereikt worden tussen leerinspanning en rendement.

Terwijl frequentielijsten op basis van ‘het’ Latijn onvoldoende representatief kunnen zijn, is het wel mogelijk om de meest frequente woorden *binnen* het pensum van een klas te laten leren/herhalen met leerwoordenlijsten op maat. Vooral in het examenjaar valt hier tijdwinst mee te behalen, tijdens de lectuur en tijdens het leren voor het SE en CE. In het tweede deel van dit artikel zal ik aan de hand van het eindexamenpensum 2021, Cicero *Pro Roscio*, ingaan op de vraag hoe het selecteren en leren van de woorden kan worden aangepakt.

### 3 Consequenties voor de lespraktijk

#### 3.1 Examenjaar

In het examenjaar moet een pensum van 4000 à 5000 woorden dermate grondig worden bestudeerd dat de kandidaat er op het CE 20 à 25 vragen over kan beantwoorden in zo weinig van de 180 minuten dat er voor een proefvertaling van 100 à 130 woorden nog genoeg tijd en zorgvuldigheid overblijft.<sup>37</sup> De inspanning om het pensum te leren varieert in semantisch opzicht sterker dan het vaste aantal van 20 pagina’s Oxfordtekst doet vermoeden. Een groot

37 Het aantal woorden van de CE-proefvertaling van de laatste vier poëzie-pensa varieert van 106 (Vergilius 2016 I) tot 125 (Ovidius 2019 II); in de laatste vier proza-pensa van 103 (Seneca 2013 I) tot 128 (Cicero 2015 II).

aantal verschillende woorden (1957 lemmata in het Livius-pensum 2017), betekent immers meer leer- of opzoekwerk dan het *Pro Roscio*-pensum, waarin maar 1348 verschillende woorden voorkomen. Bij de epische poëzie is er weliswaar sprake van een geringere totaalomvang, maar van een groot aantal verschillende woorden, zoals tabel 3 laat zien.

### 3.2 *Cicero Pro Roscio*

De hier voorgestelde maatwerkaanpak levert voor de woordenschat van het examenprogramma voor 2021 uit de *Pro Roscio* het volgende beeld op.

Van alle 188 ‘structuurwoorden’ uit de top 1000 van LASLA komen er 139 daadwerkelijk voor in de pensumselectie van de *Pro Roscio*: 71 adverbia; 28 conjuncties; 18 pronomina; 16 preposities.<sup>38</sup> Samen dekken deze 139 structuurwoorden 1978 van de in totaal 4919 woorden in het CE-pensum: 40%. De verhouding tussen het aantal te leren woorden (188) en het rendement lijkt acceptabel. Wel zijn er dan 49 woorden geleerd die in het examenjaar niet voorkomen, maar structuurwoorden leveren ook in de jaren daarvoor veel op.<sup>39</sup>

Door alle woorden die Collatinus herkent in het pensum te sorteren op frequentie en er woordsoorten aan toe te wijzen, kunnen nog eens 200 ‘inhoudswoorden’ geïdentificeerd worden die drie keer of vaker voorkomen binnen dit pensum: 82 werkwoorden, 67 substantiva en 51 adjectiva. Samen zijn deze 200 lemmata goed voor 1981 van de in totaal 4919 woorden: 39%.

Worden alle 188 structuurwoorden uit de LASLA top 1000 gecombineerd met de 200 inhoudswoorden die drie keer of vaker voorkomen, dan zou er met het leren van 388 woorden een dekkingspercentage van circa 79% worden behaald.<sup>40</sup>

Een dekkingspercentage van 78% wordt overigens ook bereikt door álle 1000 woorden uit de LASLA top 1000 te laten leren. Met de 1764 woorden van de basiswoordenlijst Latijn van Babeliowsky *et al.* wordt zelfs een dekking van 83% behaald. Maar voor deze extra dekking is wel een leerinspanning nodig die de 388 woorden uit de maatwerkaanpak aanzienlijk overtreft.

38 En zes overige woorden: *unus duo tres mille qualis tantus*. Bij deze indeling is per woord een keuze gemaakt uit dubbele mogelijkheden, immers alleen in zinsverband kan bijvoorbeeld worden vastgesteld hoe vaak *cum* als prepositie of als conjunctie is gebruikt.

39 Daarom wordt aanbevolen de lijst structuurwoorden ieder leerjaar uit te breiden en te herhalen, zodat er al voor het SE-pensum genoeg rendement is. Leren ‘voor het geval een woord in de CE-proefvertaling voorkomt’ is statistisch en pedagogisch problematisch.

40 Collatinus berekent echter een dekkingspercentage van 74%. Omdat niet alle woorden worden herkend en door verschillende attributies moet bij alle in dit artikel genoemde percentages rekening gehouden worden met een marge van onbetrouwbaarheid, naar schatting rond de 5%.

Tabel 3 Lemmata in relatie tot het totaal aantal woorden in acht examenpensa

CE	pensum	Totaal aantal woorden (x)	lemmata (y)	ratio (y/x)
2021	Cicero <i>Pro Roscio</i>	4943	1348	27%
2018	Cicero/Seneca	4571	1792	39%
2017	Livius	5030	1957	39%
2015	Cicero/Plinius	4906	1801	37%
2020	Vergilius	3938	1884	48%
2019	Ovidius	3897	1792	46%
2016	Vergilius	3838	1774	46%
2011	Vergilius	3941	1826	46%

### 3.3 Visualisatie van woordfrequentie

In de praktijk valt het niet mee om leerlingen in hun examenjaar te motiveer woorden te leren of herhalen. Door zichtbaar te maken wat het leren van woorden oplevert bij de behandeling van de tekst kan hier verandering in komen. Met Collatinus kan woordfrequentie gevisualiseerd worden in een digitale versie van een concrete tekst: woorden krijgen al naar gelang de frequentie een andere kleur. Hier zijn een paar stappen voor nodig.

Met Collatinus analyseer je eerst de frequentie van woorden binnen een concreet pensum.<sup>41</sup> Elke computeranalyse heeft als meest problematische beperking dat toeschrijvingen van losse woorden aan het juiste lemma (*eo, servus, regi*) ambigu zijn en dat alleen in context kan worden vastgesteld of bijvoorbeeld *cum* als prepositie of conjunctie geteld moet worden.<sup>42</sup> Dit is voor Collatinus niet anders. Het is echter wel mogelijk om, met aanvaardbare foutmarge, de woorden die met zekerheid drie keer of vaker voorkomen binnen een pensum te scheiden van andere.<sup>43</sup>

Aan elk van de woorden die in de  $\geq 3$  keer-groep zitten, ken je één woordsoort toe om ze in een spreadsheet op woordsoort te kunnen sorteren.<sup>44</sup> De groep woorden  $\geq 3$  keer wordt als kale lijst ingelezen en vergeleken met het tekstpensum.<sup>45</sup> Collatinus maakt dan alle herkende woorden groen in de tekst, dus de woorden die afkomstig kunnen zijn uit de ingelezen lijst van drie keer of vaker. De overige woorden, dus alle woorden die maar een of twee keer

41 Zie handleiding Collatinus, Buijckx (2019b).

42 Zie Major (2008).

43 Dus de woorden die slechts een of twee keer voorkomen of waarvan de derde instantie ambigu is.

44 Voor woorden als *cum, quod, vero* leidt dit dus tot onnauwkeurigheden in de tellingen.

45 De lijst moet ontdaan zijn van lengtetekens en één woord per regel bevatten, zie noot 29.

voorkomen, inclusief namen, blijven zwart.<sup>46</sup>

Wanneer de geleerde woorden groen worden in een projectie van de Latijnse tekst op het smartboard, ziet de leerling wat hij wel en niet geacht wordt te kennen. Het aantal woorden dat niet groen (dus niet frequent) is en dat evenmin in de examenbundels geannoteerd wordt, is doorgaans zeer beperkt. Op deze manier zien leerlingen hoe efficiënt het leren van woorden is bij de lectuur en blijft er méér aandacht over voor de zinsbouw, stijl en inhoud.

### 3.4 *Ordenen en verbanden tonen*

De structuurwoorden komen veel voor en zijn de moeite van het memoriseren waard ongeacht de feitelijke tekstkeuze in de bovenbouw. Deze woorden vijf of zes jaar lang laten leren en herhalen is ook daarom wenselijk omdat het moeilijker is ze in het geheugen te verankeren. Hun betekenissen zijn relatief abstract en verwarrend (*quisque, quidam, aliquis*), er zijn vele *false friends* (*quidam, quidem; dum, cum, tum; quisque, quoque*) en er zijn weinig behulpzame afleidingen van in de moderne vreemde talen. Wat betreft de vorm hebben vele van de meest frequente structuurwoorden slechts één of twee lettergrepen, waardoor ze bij het leren lastiger uit elkaar te houden zijn. Anderzijds zijn structuurwoorden niet verbuigbaar (behalve pronomina) en daardoor wel altijd herkenbaar in de vorm waarin ze geleerd zijn.<sup>47</sup> Het aantal woordvormingspatronen is beperkt maar kan het leren wel faciliteren: als suffix kan *-que* het meest ruime antwoord op een vraag zijn (*quisque, ubique, undique*); ook reeksen als *huc illuc* of *hinc illinc* of *tot quot, totiens quotiens* of *uter? neuter – alter – uterque* kunnen het beste in combinatie met elkaar geleerd worden.<sup>48</sup>

Voor het leren van ‘structuurwoorden’ zijn de traditionele tweetalige lijsten verdedigbaar maar is een alfabetische volgorde, alleen al voor de *qu*-woorden, verwarrend. Het groeperen van samenhangende woorden, bijvoorbeeld vraagwoorden, tijdsbepalingen en nevenschikkende voegwoorden, kan het memoriseren ondersteunen. Omdat afbeeldingen bij structuurwoorden zelden eenduidig zijn is het beter ze in korte zinnen aan te bieden: *quot homines tot sententiae; ubi bene, ibi patria*.<sup>49</sup>

Voor de inhoudswoorden is de spreiding zo groot en de overlap tussen verschillende teksten zo klein dat het rendement van jarenlang ‘vooruit leren’ be-

46 De woorden die Collatinus helemaal niet herkent (neologismen, zeer ongebruikelijke of fout gespelde woorden) worden bruin. Dit is nooit meer dan 1%. Omdat Collatinus bruine woorden niet meeneemt in de berekening van het dekkingspercentage moet een foutmarge in acht worden genomen. Namen blijven in dit artikel overigens overall buiten beschouwing.

47 Corson (1995: 172) bespreekt het *word frequency effect* in relatie tot *ease of recognition*.

48 Zie Vanheule (2015: 113).

49 Deze en andere strategieën van woorden leren worden uitgelegd op het YouTube-kanaal *ligula memoriae*, <https://www.youtube.com/channel/UCmw5brF9H1KwDpX4yVonwqQ>.



twistbaar is. Veel woorden opzoeken in het woordenboek leidt echter ook de aandacht af van de Latijnse tekst.<sup>50</sup> In een leerwoordenlijst ligt ordening per declinatie en conjugatie het meest voor de hand.<sup>51</sup>

### 3.5 Vóórdoceren

Het voorafgaand aan de lectuur behandelen – en memoriseren – van woorden die in een bepaalde passage essentieel zijn kan lectuurtempo en tekstbegrip verhogen. Al een eeuw geleden beval de Amerikaanse classicus Archibald dit aan, omdat zijn geïmmigreerde studenten *liber* niet van *library* konden afleiden. Door de belangrijkste woorden uit een tekstpassage in gesorteerde groepen van 30 vooraf te laten leren steeg het lectuurtempo aanzienlijk.<sup>52</sup>

Zo kan ook de aandacht gericht worden op de semantische waarde en vorming van de Latijnse woorden, zoals het verschil tussen *turba*, *vulgus*, *multitudo* of de adjectivische afleiding op *-eus* van *caelum* bij *caeruleus*. Bovendien kunnen verbanden met moderne vreemde talen, woordvormingsprincipes, contextzinnen (*Sicilia insula est*) en afbeeldingen mnemotechnische ondersteuning bieden, zoals in de zeer creatief vormgegeven Bamberger Wortschatz (zie afb. 2). Het doorzien van de samenhang tussen Latijnse woorden onderling, zoals in het *Pegasus Vocabularium*, kan leiden tot meer aandacht voor de Latijnse woorden zelf in plaats van alleen voor de gegeven betekenissen in een woordenlijst.<sup>53</sup>

Een andere manier van vóórdoceren is een PowerPoint tonen van de woorden die in een bepaalde passage cruciaal zijn. Per woord wordt de klank hardop uitgesproken, worden er plaatjes getoond die als *mnemonic* helpen, en verschijnen er achtereenvolgens verwante Latijnse woorden, een contextrijk zinnetje waarin het woord zichzelf illustreert en woorden uit moderne vreemde talen die van het Latijnse woord zijn afgeleid. Bij extra moeilijk te onthouden woorden kunnen humoristische *keyword*-zinnnetjes helpen om

50 Wanneer woorden voornamelijk opgezocht worden in een alfabetische woordenlijst achterin het boek, waar veel minder woorden staan en vaak slechts een of twee betekenissen, dan wordt er geen Latijn lezen maar bladeren ingeoeft en is de overgang naar 'Pinkster' nog groter.

51 Zie Lenaers (1997).

52 Bijvoorbeeld twintig neutra van de 3<sup>de</sup> declinatie en tien verba van de 2<sup>de</sup> conjugatie, Archibald (1914: 267).

53 Door deze woorden systematisch te groeperen (per declinatie/conjugatie) wordt het herkennen in morfologische variatie voorbereid. Er worden dan vaste woordvormingspatronen zichtbaar: *-tas* en *-tudo* vooral bij substantiva uit de 3<sup>de</sup> declinatie; *-bilis* vooral bij adjectiva die van werkwoorden gemaakt zijn en die verbogen worden volgens de 3<sup>de</sup> declinatie; *-eus* bij adjectiva die een materiële eigenschap weergeven en altijd verbogen worden volgens de 1<sup>ste</sup> declinatie, zie ook Lenaers (1997) en Ackerman (2014). De Adeo-Norm lijst van Utz (zie afb. 2) maakt zichtbaar hoeveel hulp de leerling hier bij het memoriseren aan kan ontnemen. Voor Utz' verantwoording van de *Bamberger Wortschatz*, zie [https://www.ccbuchner.de/titel-1-1/adeo\\_norm-549/](https://www.ccbuchner.de/titel-1-1/adeo_norm-549/).



klankbeeld en betekenis visueel te verankeren.<sup>54</sup> Het ‘googelen’ van de juiste afbeelding kost veel tijd en het uitbesteden hiervan aan leerlingen leidt tot plaatjes van een kapsel bij het woord *illam*. Maar het delen van dit maatwerk tussen collega’s, in de vorm van lijsten, PowerPoints of overhoorapplicaties kan veel werk besparen.<sup>55</sup>

Woordenlijsten in digitale vorm kunnen leerlingen ook importeren in hun favoriete applicaties als Quizlet en Memrise waarin ook plaatjes, contextzinnen en ingesproken audio kunnen worden ingevoerd. Op deze manier kan het leren van woorden een audiovisuele, contextrijke activiteit worden in plaats van het ‘stampen’ van losse woorden die niet vaak genoeg terugkomen in teksten om ooit het langetermijngeheugen te kunnen bereiken.

Een van de meest effectieve en ludieke manieren van het vóórdoceren van woorden is het gericht inzetten van ‘levend Latijn’ als activerende werkvorm.<sup>56</sup> Het is daarbij belangrijk dat de leerling uit de context kan begrijpen wat nieuwe woorden betekenen: *comprehensible input*. Door een mes te tonen kunnen in een vraag- en antwoordspel woorden blijvend in het langetermijngeheugen verankerd worden, omdat de klas ze in korte, voorgestructureerde antwoorden herhaalt.<sup>57</sup> De klas hoeft dus zelf geen vloeiend Latijn te spreken en de docent kan zijn voorbereide vragen en opdrachten oplezen. *Total Physical Response* (TPR) werkt systematisch met eenvoudige opdrachten zodat de leerlingen door die uit te voeren laten zien dat ze Latijn begrijpen (*Saluta vicinum! Aperi oculos ambos! Iacite libros ad magistrum!*). Ze horen de taal in zogenaamde ‘chunks’ (betekenisvolle combinaties) waardoor niet meer elk woord in principe alles zou kunnen zijn, maar in een bepaalde context betekenis krijgt.<sup>58</sup>

Het voordeel van actief Latijn als werkvorm is daarmee dat leerlingen taal verwerven in plaats van moeizaam stampen, op een manier die blijvend effect heeft en de tussenstap van vertalingen naar het Nederlands vervangt door het zichtbaar maken of de input correct begrepen is.<sup>59</sup>

54 Een keyword-zinnetje koppelt de betekenis van een woord aan de klank ervan, bijvoorbeeld voor caput: ‘Kaputt ist der Kopf’, zie Geisselhart en Lange (2014); Shapiro en Waters (2005); Webb en Nation (2017: 115-6).

55 Voor inzage in de beeldbank *ligula memoriae* kunt u een email sturen aan mbuijck@hetccc.nl. YouTube-kanaal *ligula memoriae* bevat diverse strategieën om woorden blijvend te verankeren in het geheugen, zoals *keyword*- en contextzinnen en woordafleiding.

56 Zie Porton (2018) en Regtuit (2015).

57 ‘*Quid est?*’ ‘*sica est*’ ‘*Estne sica gladius parvus an magnus?*’ ‘*Estne Roscius sicarius an latro?*’. Zie ook ‘Lecturer of the Year Election RUG 2015: Remco Regtuit’, <https://youtu.be/5ULmmQdrLG8>. Saffire (2006) bevat uitgeschreven scripts hoe dit voor het Grieks werkt nog vóór het alfabet is aangeleerd.

58 Zie <https://www.mezzoguild.com/chunking-and-language-learning/>.

59 Cursussen van Addisco.nl en Polisjerusalem.org zijn mede gebaseerd op Asher (2012). De leerboeken *Forum* en *Polis* van Christophe Rico werken voornamelijk met TPR.

### 3.6 Naar een nieuwe basiswoordenlijst?

Of het zinvol is om een nieuwe basiswoordenlijst uit te geven is de vraag. Enerzijds verklaart het ontbreken van afspraken over woordenschat ten dele de problemen op het examen. Bij Latijn is ‘Pinkster’ het meest gebruikte woordenboek. Omdat alle woorden daarin staan, wordt er bij de gelezen teksten, in tegenstelling tot het CE Grieks, niets geannoteerd. In de acht geanalyseerde pensum gaat het om 1300 à 1900 unieke woorden (zie tabel 3). De kandidaten moeten ofwel zelfs de meest ongebruikelijke woorden (in tekstverband) leren, ofwel kostbare examentijd besteden aan bladeren in het woordenboek. Als er consensus zou zijn over een basiswoordenlijst, zoals over de grammatica-minimumlijst, dan zou het mogelijk zijn voortaan te annoteren wat daar niet in staat.

Anderzijds verklaart de wet van Zipf mede waarom zo’n consensus er nooit is gekomen: de vrijheid van docenten bij het SE-pensum en het jaarlijks wisselen van het CE-pensum maken een gemeenschappelijke basiswoordenschat onmogelijk. Zelfs als een uitgever een frequentielijst zou baseren op alle teksten in zijn eigen bloemlezing, zal geen leerling er meer dan 30 pagina’s Oxfordtekst uit lezen. Onder de huidige omstandigheden en gelet op de inherente wetmatigheden van woordfrequentie, vooral bij de inhoudswoorden, lijkt een combinatie van vaste en flexibele lijsten het maximaal haalbare. Er zou bijvoorbeeld geëxperimenteerd kunnen worden met:

- Een vaste lijst van *structuurwoorden* (zoals de 188 uit de LASLA top 1000), eventueel aangevuld met onderling samenhangende woorden als *uter? uterque neuter alter*. De lijst kan per woordsoort worden aangeboden en vervolgens systematisch onderverdeeld, bijvoorbeeld: bijwoorden van tijd en ruimte, onder- en nevenschikkende voegwoorden, voorzetsels met een bepaalde naamval.<sup>60</sup>
- Een vaste lijst van *inhoudswoorden met algemene betekenis* (circa 100?). Bij zeer algemene woorden als *res* helpen enkele korte contextzinnen of collocaties (*res publica, rem gerere*) méér dan een lijstje vertalingen.<sup>61</sup> Door van *petere* en *agere* twee of drie basisbetekenissen te laten leren, kan ook beslist worden of er bijzondere betekenissen geannoteerd moet worden in tekstuitgaven en bij toetsen.
- Een vaste lijst van *cultuurgebonden woorden* (50?), waarvan niet zozeer de vertaling in het Nederlands het probleem is (*consul, senator*) maar het concept (*pietas, iudex*) en die de integratie tussen taal en cultuur versterken. Naar het voorbeeld van *Pegasus Vocabularium* zijn hier geen betekenissen maar omschrijvingen gewenst.<sup>62</sup>

60 Zie Brouwer (1991). Na het memoriseren van de acht voorzetsels die met de ablativus gaan (*ab, ex, in, de, pro, cum, sine, prae*) kan de leerling gokken dat na andere voorzetsels een accusativus komt. *In* met de accusativus heeft een andere betekeniswaarde en komt vaak genoeg voor om het verschil te laten leren, bij *sub* en *super* is er geen verschil in vertaling.

61 Over collocaties, zie Peters (2017: 45-61).

62 ‘Quaestor. De *quaestores* waren verantwoordelijk voor de financiële administratie en voor de schatkist’ (Ackerman 2018: 47).

- Een uitgebreide lijst van principes van *woordvorming*: prefixen en suffixen en de vaste patronen waarmee woorden gevormd kunnen worden uit één stam. De voorbeelden moeten zo gekozen worden dat die woorden niet alleen exemplarisch zijn maar ook verwantschap met de moderne vreemde talen inzichtelijk maken (*horrēre* + *-bilis* => *horrible*). Enerzijds helpt dit bij het memoriseren van de Latijnse woorden, anderzijds leert de leerling zo hoe hij zijn investeringen bij de klassieke talen in het *academic English* van zijn vervolgstudie kan inzetten.<sup>63</sup>

Deze vier vaste onderdelen zouden in de leergangen verwerkt kunnen worden en regelmatig herhaald; Vlaamse leergangen als *Tolle Lege* en *Pegasus* beginnen vaak met een blok woorden: ‘Ken je deze nog?’. Het activeren van voorkennis is immers een belangrijke voorwaarde om tot tekstbegrip te komen.<sup>64</sup>

Daarnaast zijn er in ieder leerjaar *flexibele aanvullingen* nodig: een katern of digitaal bestand met alle inhoudswoorden die in een bepaald pensum drie keer of vaker voorkomen.<sup>65</sup> In feite ligt hier hetzelfde principe aan ten grondslag als aan de minimumlijst stilistische middelen die per genre eigen aanvullingen heeft. Alleen zal in dit geval blijken dat de gemeenschappelijke deler van leerwoorden uit een epos-examen of een proza-examen zéér gering is.<sup>66</sup> Wanneer classici de vertaalopgave van het centraal examen bekijken, baseren zij hun oordeel vaak op syntaxis of morfologie en constateren dat er in dat opzicht steeds makkelijker teksten worden gekozen. Het zou de moeite waard zijn om te onderzoeken of, voor leerlingen, veeleer de semantiek van de proefvertaling de verklarende factor zou kunnen zijn voor slechte examenresultaten, zeker bij poëtisch taalgebruik.

#### 4 Conclusie

Niet langer te verdedigen zijn drie uitgangspunten die in de vorige eeuw breed gedeeld werden. Ten eerste dat er in de bovenbouw geen tijd meer besteed zou hoeven worden aan consolidatie en uitbreiding van woordenschat en grammatica: leerpsychologisch onderzoek naar de ‘vergeetcurve’ is kennelijk in verge-

- 63 Corson (1995) schat dat 60% van de woorden uit het *academic English* uit het Grieks of Latijn afkomstig zijn. Corson (1997) bespreekt welke leerpsychologische en sociologische hindernissen moeten worden overwonnen bij het memoriseren van juist deze woorden. Veel aandacht voor woordvormingspatronen heeft Lenaers (1997).
- 64 Van der Plaat (2015) geeft goede adviezen welk type woorden op welke manier aangeboden kan worden.
- 65 Op [stilus.nl/woorden](http://stilus.nl/woorden) worden dergelijke lijsten gepubliceerd, voor *Pro Roscio* ook in Word, Quizlet, Memrise.
- 66 Zelfs tussen het ene of het andere Livius-pensum is de overlap beperkt: een jaar lang teksten lezen over Romeinse religie bereidt nauwelijks voor op een proefvertaling over Hannibals strategie (CE 2017).

telheid geraakt.<sup>67</sup> Ook is het helaas niet mogelijk om een ‘dekkingsgraad’ van 85% te bereiken met basiswoordenlijsten, een percentage dat sowieso te laag ligt om onbekende woorden uit de context te kunnen afleiden. Het meest problematisch is ten slotte de aanname dat het leren van een lijst woorden die in ‘het’ Latijn het meest frequent zouden zijn ook maar enige garantie kan bieden voor een zeer beperkt maar onvoorspelbaar pensum authentieke teksten.<sup>68</sup> Al lang voor de uitvinding van de computer had onderzoek van statistici en taalkundigen immers de uitgangspunten ter discussie gesteld onder een basiswoordenlijst bij een minieme steekproef van 50 pagina’s Oxfordtekst. Het goede nieuws is dat dankzij gebruiksvriendelijke en gratis software elke classicus voor zijn eigen klassen en leatuurkeuze zichtbaar kan maken wat het rendement is van de geleerde woordenschat in een concreet pensum. Zo wordt woorden leren niet per se leuker of makkelijker, maar wel minder zinloos.<sup>69</sup>

Orfeo gymnasium van het Carolus Clusius College  
Kamperweg 1  
8019 AV Zwolle  
m.buijkx@hetccc.nl

## Websites

Addisco: <http://www.addisco.nl>  
Stilus: <http://www.stilus.nl/woorden>  
Collatinus: <https://outils.bibliissima.fr/en/collatinus/>  
LASLA: <http://web.philo.ulg.ac.be/lasla/>  
Perseus under Philologic: <http://perseus.uchicago.edu/>  
Hulstaert: <https://www.kdg.be/latijn-lezen>  
Janus: <https://janus.humanities.uva.nl/>  
Pinkster: <http://www.latijnnederlands.nl/>

## Bibliografie

Ackerman, J., W. Cool, M. Hillewaere, B. Mertens, W. van der Meiren, J. Van Eetvelde en A. Verlingen. 2018. *Pegasus Vocabulary*, Kalmthout.  
Amerongen, R. van, W. Budé, H. Derix, H. van Gessel, T. Goossens, W. Mak, Ph. Roorda en J. Zwart. 1986. ‘Leerplan Latijn’, *VCN bulletin* 11, 1-39. (<https://www.stilus.nl/examen/x-0124.pdf>)  
Archibald, H.T. 1914. ‘Latin Vocabulary and the English Derivative’, *The Classical Journal* 9, 265-271.

67 Over leerpsychologische aspecten zie Vanheule (2015); Carlon (2016); Corson (1997); Shapiro en Waters (2005); Vis (2013). De vergeetcurve is ontdekt door Ebbinghaus in 1885.

68 Bull (1949: 469): ‘any word count is a statistically valid report only on what is included in it.’

69 Zeer hartelijk dank ik de redactie van *Lampas* en de collega’s die dit artikel door middel van hun waardevolle feedback veel zinvoller hebben gemaakt.

- Asher, J. 2012. *Learning Another Language Through Actions*, Los Gatos.
- Babeliowsky, J.K.L., D. den Hengst, W. Holtland, W. van Lakwijk, J. Th. K. Marcellis, H. Pinkster, J.J.L. Smolenaars. 1975. *Basiswoordenlijst Latijn*, 's Gravenhage.
- Blanchard, D., F. Cerruti, M. Kopf, S. Morassut, C. Rico en N. Schonebaum. 2015. *Forum. Speaking Latin as a living language*, Jeruzalem.
- Brouwer, H., A. van den Heuvel, J Korteweg en J. Moerbeek. 1991. *Index. Basiswoorden Latijn geordend naar themavelden*, Emmeloord.
- Buijck, M.H. 2019a. 'Woordenschat. Maatwerk met Collatinus', *VCN-bulletin* 44, 20-21.
- Buijck, M.H. 2019b. 'Handleiding Collatinus', <https://www.stilus.nl/woorden/mb-collatinus.docx>.
- Bull, W.E. 1949. 'Natural Frequency and Word Counts', *The Classical Journal* 44, 469-484.
- Carlson, J. 2016. 'Quomodo dicitur? The importance of memory in language learning', *Teaching Classical Languages* 7, 109-135.
- Corson, D. 1995. *Using English Words*, Dordrecht.
- Corson, D. 1997. 'The Learning and Use of Academic English Words', *Language Learning* 47, 671-718.
- Delatte, L., E. Evraert, S. Govaerts en J. Denooz. 1981. *Dictionnaire frequentiel et index inverse de la langue latine*, Luik. (<https://www.stilus.nl/woorden/lasla1981.pdf>)
- Diederich, P. 1939. *The Frequency of Latin Words and Their Endings*, Chicago. (<https://www.stilus.nl/woorden/diederich.pdf>)
- Geisselhart, O. en H. Lange. 2014. 'Kaputt ist der Kopf', München. (<https://www.m-vg.de/mediafiles/Leseprobe/9783868825299.pdf>)
- Gruber-Miller, J. (ed.). 2006. *When Dead Tongues Speak. Teaching beginning Greek and Latin*, Oxford.
- Hu, M. en I. Nation. 2000. 'Unknown Vocabulary Density and Reading Comprehension', *Reading in a Foreign Language* 13, 403-430.
- Hulstaert, K. 2016. *Latijn lezen, zien begrijpen. De positionele methode in de klas*, Gent.
- Hupperts, C. en J. de Jong. 1998. *Vox*, Leeuwarden.
- Jansen, M. en A. Jaspers. 2017. 'Wet van Zipf in taal verklaard'. <https://www.nemokennislink.nl/publicaties/wet-van-zipf-in-taal-verklaard/>.
- Kanters, H.J. 1979. *Basiswoorden Latijn*, Groningen.
- Lenaers, R. 1997. *Verba. Latijns basisvocabularium*, Deurne.
- Lestrade, S. 2017. 'Unzipping Zipf's law', *PLoS ONE* 12.8, e0181987.
- Major, W.E. 2008. 'It's Not the Size, It's the Frequency. The value of using a core vocabulary in beginning and intermediate Greek', *Committee for the promotion of Latin Online*, Winter 2008. (<https://camws.org/cpl/cplonline/files/Majorcplonline.pdf>)
- Nation, I.S.P. 2013. *Learning Vocabulary in Another Language*, Cambridge.
- Nellissen, L. 2018. 'Stilus-woordenlijst op frequentie'. (<https://www.stilus.nl/woorden/stilusfrequentielijstlatijn.xlsx>)
- Peters, E. 2017. *Woordenschat aanleren in een vreemde taal. Hoe doe je dat?*, Leuven en Den Haag.
- Plaat, A. van der. 2015. 'Woordenschatverweving Grieks (en Latijn). De optimalisering van het vocabulaire met het oog op het onderwijs in het Grieks', *Lampas* 48, 265-277.
- Porton, C. 'Actief Latijn als krachtig didactisch middel', <https://www.addisco.nl/actief-latijn-als-krachtig-didactisch-middel/>.
- Porton, C. 2018. 'De klassieke talen op de rand van een (r)evolutie', *Kleio* 47, 32-51.
- Regtuit, R.F. 2015. 'Latijn is een taal, geen puzzel. Het gebruik van actief Latijn in het onderwijs', *Lampas* 48, 286-297.
- Rico, C. 2015. *Polis. Speaking ancient Greek as a living language*, Jeruzalem.
- Saffire, P. 2006. 'Ancient Greek in Classroom Conversation' in J. Gruber-Miller (ed.), 158-189.
- Shapiro A.M. en D.L. Waters. 2005. 'An Investigation of the Cognitive Processes Underlying the Keyword Method of Foreign Vocabulary Learning', *Language Teaching Research* 9, 129-146.
- Utz, C. 2001. *Adeo Norm mit Lernhilfen. Das Lateinische Basisvokabular*, Bamberg.
- Vanheule, J. 2015. *Latijnse woorden: blokken of inoefenen? Een onderzoek naar de effectiviteit en de*

*evaluatie van de paired-associate-methode en de involvement-load-methode bij Latijnse woordenschatverwerving en -berhaling*, Masterscriptie Universiteit Gent.

(<https://lib.ugent.be/nl/catalog/rug01:002213101>)

Vis, J. 2013. 'De verwerving en didactiek van het Oudgriekse vocabulaire', *Lampas* 46, 222-233.

Webb, S. en P. Nation. 2017. *How Vocabulary is Learned*, Oxford.

Wellink, J.G.J.M. 1997. *Basisvocabularium Latijn*, Deurne.

Williams, M.A.E. 2012. *Essential Latin Vocabulary*, Californië.

Zipf, G.K. 1949. *Human Behaviour and the Principle of Least Effort. An introduction to human ecology*, Cambridge.



53e jaargang nummer 3

oktober 2020



# AMPAS

tijdschrift voor classici

JOHAN WESTSTEIJN	Een Arabische Orestes	298
EMILY HEMELRIJK	Op weg naar vrijheid en burgerschap Beelden van vrouwelijke vrijgelatenen	319
BRAM FAUCONNIER	Achter de schermen van de festivalwereld De verenigingen van atleten en artiesten in de Keizertijd	342
MICHAEL BUIJKX	Woordenschatgraven Over frequentielijsten en vocabulaireverwerving	360
EVELIEN BRACKE	Didactische bijdrage De impact van Oudgriekse studie op kansarme leerlingen Participatieverruiming in het lagere scholenproject Oude Grieken – Jonge Helden	384