

Pieterjan Bonne (a), Joke Vrijders (a) & Jordi Casteleyn (b)

(a) Arteveldehogeschool, Gent

(b) Universiteit Antwerpen

Contact: [Pieterjan.bonne@arteveldehs.be](mailto:Pieterjan.bonne@arteveldehs.be)

[Joke.vrijders@arteveldehs.be](mailto:Joke.vrijders@arteveldehs.be)

## Stap voor stap, woord voor woord – op zoek naar een kern van academische woordenschat

### 1. Inleiding

Bij de overgang naar het hoger onderwijs komen studenten in contact met academische taalvaardigheid. Een aspect van deze academische taalvaardigheid is woordenschat (Cajot, Heeren & Vrijders 2018). Academische woordenschat is tweeledig. Naast echte vaktaal – bijvoorbeeld *gameten* en *zygoten* voor een student Biologie – is er ook het algemeen academisch Nederlands dat over verschillende vakdomeinen wordt gebruikt, bijvoorbeeld *hypothese* of *relateren*.

Academische woordenschat is een belangrijke sleutel om teksten te begrijpen of lessen te kunnen volgen. Voor tekstbegrip moet volgens onderzoek 95% van de woorden gekend zijn; voor gedetailleerd begrip zelfs 98% (Laufer & Ravenhorst-Kalovski 2010; Nation 2006).

Om (toekomstige) studenten te ondersteunen bij de start in het hoger onderwijs, hebben veel scholen uit het secundair of voortgezet onderwijs, hogescholen en universiteiten aandacht voor algemeen academische woordenschat in het Nederlands. Ze doen dat onder andere met woordenlijsten en oefeningen. Een bekend naslagwerk met een woordenlijst is *Wijze Woorden* (Giezenaar & Schouten 2002).

Veel scholen en instellingen hebben ook hun eigen lijsten. Als basis om die lijsten te maken, gebruikt men soms een eigen corpus, maar die zijn vaak beperkt tot geschreven materiaal of materiaal uit de eigen instelling. Zo is *Wijze Woorden* gebaseerd op een corpus met materialen uit wetenschapsbijlagen van dagbladen en uit studiemateriaal van een groot aantal universitaire studierichtingen aan een Noord-Nederlandse universiteit.

Er bestaat geen algemeen gesproken en geschreven corpus academisch Nederlands, waardoor er geen stevige basis is om uitspraken te doen over welke woorden de moeite zijn om te leren. Deze bijdrage wil een eerste stap zetten in de zoektocht naar een

kern van algemene academische woordenschat. Als verkenning verzamelden en analyseerden we 35 verschillende bronnen met een focus op het ondersteunen van (toekomstige) studenten bij de verwerving van algemene academische woordenschat. Na een korte toelichting over de woordenlijsten zelf, geven we een korte analyse van de resultaten. Welke woorden zijn de moeite om te leren?

## 2. De woordenlijsten

Van maart tot juni 2018 deden we verschillende oproepen om bronmateriaal rond woordenschat academisch Nederlands te delen. De oproepen werden verspreid via het Nederlands/Vlaams Platform Taalbeleid Hoger Onderwijs, het Netwerk Didactiek Nederlands en het Forum Taalbeleid Hoger Onderwijs. We ontvingen 30 antwoorden met documenten of verwijzingen naar – soms dezelfde – materialen. Deze vulden we aan met websites en naslagwerken rond overgang naar het hoger onderwijs en academisch taalgebruik. We kwamen tot 35 verschillende bronnen.

De 35 bronnen kwamen zowel uit het hoger ( $n = 28$ ) als het secundair onderwijs ( $n = 5$ ), en zowel uit Vlaanderen ( $n = 25$ ) als uit Nederland ( $n = 9$ ). De bronnen richtten zich echter niet per definitie op een niveau of een land. Een Vlaamse bron uit het hoger onderwijs kon ook gebruikt worden in een Nederlandse school voor voortgezet onderwijs, bijvoorbeeld. Slechts drie bronnen richtten zich specifiek op het hoger onderwijs.

De bronnen waren, naast woordenlijsten, ook oefeningenbundels of -websites, lesmaterialen of hoofdstukken uit boeken. Uit deze bronnen selecteerden we de woorden die als voorbeeld van academische woordenschat werden gegeven, die als antwoord of antwoordsuggestie op multiplechoicevragen werden meegegeven, die waren gearceerd of gemarkeerd als academische woordenschat in teksten, enz. Op die manier werd elke bron in een woordenlijst omgezet.

De lengte van de woordenlijsten verschilde. De kortste lijst telde slechts 12 (instructie) woorden; de langste telde 1.472 woorden. De gemiddelde lengte van de lijsten is 324 woorden, de mediaan ligt op 258.

## 3. Op zoek naar een gedeelde kern

Om richting een gedeelde kern te kunnen gaan, zijn we vertrokken van deze woordenlijsten als een corpus. We combineerden alle woordenlijsten tot een omvattende woordenlijst. Deze lijst bevatte 11.976 zogeheten *tokens* (= alle woorden of woordcombinaties, dus inclusief dubbels, verbogen of vervoegde vormen). Hiervan waren er 7.463 uniek of zogeheten *types*. 1.113 ervan waren collocaties of woordcombinaties.

We brachten alle woorden en woordcombinaties onder in woordfamilies. Een woordfamilie bevat het woord en alle afleidingen van dat woord. Bijvoorbeeld: *resultaat-resulteren-resultante*. In de lijst staan 5.875 woordfamilies.

De woordenlijst werd op twee manieren geanalyseerd. Eerst keken we wat de overlap was tussen de verschillende bronnen in de woordenlijst. Vervolgens bekeken we welke types en woordfamilies overlaptten. Hiermee vormden we een eerste gedeelde kern.

Uit analyse van de woordenlijst bleek dat er weinig overlap is in types en woordfamilies. Er is dus slechts een beperkte gedeelde kern over de verschillende bronnen. We keken hiervoor zowel naar het totale aantal keren dat een woord voorkomt als naar het aantal keren dat een woord voorkomt in een specifieke bron. Voor het totale aantal keren dat een woord voorkomt, werd meer of minder dan 10 als grens genomen. Voor het aantal keren dat een woord voorkomt in een specifieke bron werd in eerste instantie gekeken naar een overlap van 50% (meer dan 18 bronnen), maar geen enkel type of geen enkele woordfamilie bleek in minstens de helft van de bronnen voor te komen. Vervolgens werd dit bijgesteld naar minstens 30% van de bronnen (meer dan 10 bronnen).

Tabel 1 geeft een overzicht van de analyses op ‘voorkomen op type’ en ‘voorkomen op woordfamilie’. Van de types komt ongeveer 75% slechts 1 keer of in 1 bron voor. Slechts 30 woorden (0,40%) komen meer dan 10 keer voor, waarvan sommige meermaals in dezelfde bron. Slechts 10 woorden (0,05%) komen in meer dan 10 bronnen voor. *Hypothese* komt in het meeste aantal verschillende bronnen voor. Het komt in de lijst 16 keer voor in 15 verschillende bronnen (43%). Van de woordfamilies komt ongeveer 66% amper 1 keer of in 1 bron voor. 116 woordfamilies (1,97%) komen meer dan 10 keer voor, waarvan sommige in verschillende vormen in dezelfde bron. 33 woordfamilies (0,56%) komen in meer dan 10 bronnen voor. De verschillende vormen van de woordfamilie *relateren-relatie-gerelateerd-relatief-relationeel* komen het meest voor: 29 keer in de lijst in 13 verschillende bronnen. *Hypothese-hypothetisch* komt in het meeste aantal verschillende bronnen voor. In 16 verschillende bronnen (46%) komen de twee vormen samen 20 keer voor.

	<b>komt slechts 1 keer voor</b>	<b>komt in slechts 1 bron voor</b>	<b>komt meer dan 10 keer voor</b>	<b>komt in meer dan 10 bronnen voor</b>
Types	73,48%	74,88%	0,40%	0,05%
Woordfamilies	66,72%	70,94%	1,97%	0,56%

Tabel 1 – Overzicht analyses voorkomen op type en woordfamilie.

Samenvattend zien we dat de types en de woordfamilies voor het leeuwendeel in slechts één bron voorkomen. Nog geen procent van de types of woordfamilies komt in min-

stens 30% van de bronnen voor en geen enkel type of geen enkele woordfamilie komt in meer dan de helft van de bronnen voor.

Om te kijken welke types en woordfamilies overlappen, en zo de gedeelde kern in kaart te brengen, werkten we op bron-niveau en keken we naar zowel type als woordfamilie. We werkten op bron-niveau omdat sommige bronnen meerdere keren hetzelfde woord aanbrenge(n), bijvoorbeeld in oefeningen, instructies en teksten. Dat geeft een vertekend beeld van het belang. Het is waardevol om naar woordfamilies te kijken, omdat de vorm waarvoor gekozen wordt niet noodzakelijk de vorm is die het meest of als enige voorkomt of gebruikt wordt. Naast woordfamilies vinden we ook type interessant. Het is immers niet altijd evident dat een student van een woord alle afgeleiden kent of herkent.

De grens van voorkomen die we hebben getrokken in de analyses – in meer dan 30% van de bronnen – is arbitrair. Voor een breder overzicht namen we ook de types en woordfamilies op die in slechts 10 verschillende bronnen voorkwamen. Zo bestaat het overzicht uit 21 types en 50 woordfamilies. Meer gedetailleerde informatie hierover is te vinden op: [www.arteveldehogeschool.be/taal](http://www.arteveldehogeschool.be/taal) bij ‘Tools voor docenten’.

#### 4. Conclusie

Er circuleren veel materialen voor academische woordenschat in het secundair of voortgezet onderwijs en in het hoger onderwijs. Deze zijn telkens pogingen van betrokken onderzoekers, docenten en taalbeleidsmedewerkers om de woordenschat van het hoger onderwijs te vatten en door te geven. Welke lijst nu het beste het academisch Nederlands weerspiegelt, is niet te bepalen. Ook is het lastig vaststellen welke lijsten wel of niet betrouwbaar zijn.

Wat we wel kunnen vaststellen, is dat de academische woordenschat die je als (toekomstig) student voor ogen krijgt in deze materialen enorm verschilt. Er is slechts beperkte overlap tussen een beperkt aantal lijsten.

Om verder zicht te krijgen op wat (toekomstige) studenten echt nodig hebben, moet verder werk gemaakt worden van een corpus gesproken en geschreven academisch Nederlands.

#### Referenties

Cajot, G., J. Heeren & J. Vrijders (2018). “Academische taalvaardigheid in het hoger onderwijs: Een sleutel voor een succesvolle overgang”. Mondelinge presentatie voor de VLOR, Brussel. Online raadpleegbaar op: <https://assets.vlor.be/www>.

[vlor.be/attachment/Academische%20Taalvaardigheid\\_Presentatie%20VLOR%20241018\\_0.pdf](http://vlor.be/attachment/Academische%20Taalvaardigheid_Presentatie%20VLOR%20241018_0.pdf).

Giezenaar, G. & E. Schouten (2002). *Wijze woorden. Woordenlijst academisch Nederlands met idioom oefeningen*. Amsterdam/Antwerpen: Intertaal.

Laufer, B. & G.C. Ravenhorst-Kalovski (2010). "Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension". In: *Reading in a foreign language*, 22 (1), p. 15-30.

Nation, I. (2006). "How large a vocabulary is needed for reading and listening?". In: *Canadian modern language review*, 63 (1), p. 59-82.