

Zoek de fout; een foutenclassificatie als aanzet tot gerichte remediëring Nederlands in het hoger professioneel onderwijs in Vlaanderen

ANNELIES DEVENEYNS & JOSÉ TUMMERS

Leraren in Vlaanderen zijn het erover eens dat de schriftelijke taalvaardigheid van jongeren achteruitgaat. Ook in het hoger onderwijs stelt men vast dat studenten steeds minder in staat zijn een vormelijk correcte tekst te schrijven. Desondanks is er relatief weinig onderzoek verricht naar de schriftelijke taalbeheersing Nederlands van studenten en blijven spelling en woordenschat de primaire aandachtspunten in het aanbod Nederlandse communicatie. Deze verkennende studie heeft tot doelstelling om in kaart te brengen welke de meest typische en frequente fouten zijn die studenten in het hoger professioneel onderwijs maken en wat dat voor de onderwijspraktijk betekent. Een foutenanalyse van teksten geschreven door eerstejaars bachelorstudenten toonde aan dat spelling – hoewel vaak de focus van remediëringstrajecten – niet het grootste probleem vormt. Teksten van studenten worden vooral ontwricht door een foutief gebruik van taalmarkeerders. Die vaststelling kan een aanzet zijn tot een gerichte remediëring van de schriftelijke taalvaardigheid bij studenten.

In het hoger onderwijs in Vlaanderen leeft de overtuiging dat het taalgebruik, vooral het schriftelijk taalgebruik, van studenten moet

worden bijgespijkerd. Ook in de algemene media wordt er over dit thema regelmatig aan de alarmbel getrokken, met onder andere de terugkerende taalbijlagen, bijvoorbeeld in *De Standaard* (23-27/10/2012), waar het gros van de aandacht doorgaans spelling en woordenschat te beurt valt. Het schaarse wetenschappelijk onderzoek dat voor Vlaanderen voorhanden is, weerlegt de overwegend negatieve perceptie niet (Bogaert & Verheyden, 2010; Peeters & Van Houtven, 2010; Van Houtven et al., 2010). Ter verklaring wordt gewezen op de kloof tussen de eindtermen van het secundair onderwijs en de – impliciete – taalvereisten in het hoger onderwijs waar een abstracter en formeler taalgebruik gehanteerd wordt (Van den Branden, 2010).

Ondanks de prominente positie die taal tijdens het laatste decennium in het Vlaamse onderwijsbeleid inneemt (Vandenbroucke, 2007; Smet, 2011), laat het niveau van de (schriftelijke) taalbeheersing in het hoger onderwijs nog steeds te wensen over. We stellen een toename vast van initiatieven om het taalniveau van studenten bij hun instap in het hoger onderwijs te peilen (Deygers & Kanobana, 2010; De Wachter & Heeren, 2011, 2013) en nadien door middel van remedië-

ringstrajecten op peil te brengen (Sterckx & Vanhoren, 2011).

Rekening houdend met de beperkte middelen die beschikbaar zijn voor taalzorg en -remediëring in het hoger onderwijs lijkt het ons nuttig om het probleem duidelijk te omschrijven. Daartoe willen we in deze bijdrage een aanzet geven. We brengen de fouten in kaart die studenten in professionele bacheloropleidingen maken, aangevuld met hun frequentie en verspreiding. De resulterende foutenclassificatie kan als leidraad fungeren bij het bepalen van prioritaire aandachtspunten in de taalremediëring voor professionele bachelorstudenten.

Methode

Steekproefsamenstelling en opdracht

Deze foutenclassificatie maakt deel uit van een onderzoeksproject naar de schriftelijke taalbeheersing van professionele bachelorstudenten aan de Katholieke Hogeschool Leuven (KHLeuven). De KHLeuven is een Vlaamse hogeschool met 13 professionele bacheloropleidingen in de volgende domeinen: handelswetenschappen en bedrijfskunde, lerarenopleiding, sociale wetenschappen, gezondheidszorg en technologie. Binnen het Europees Kwalificatiekader beantwoorden die opleidingen aan de vereisten van niveau 6.¹ In vergelijking met academische programma's aan universiteiten en hogescholen vormt de studentenpopulatie van professionele bacheloropleidingen een getrouwere dwarsdoorsnede van de uitstroom uit het secundair onderwijs, doordat er naast studenten met een aso-achtergrond ook aanzienlijke groepen jongeren met een tso- en bso-diploma studeren.^{2, 3}

Uit een initiële clustersteekproef van 346 teksten over de 13 opleidingen is er, rekening houdend met de arbeidsintensiviteit van de manuele foutenidentificatie en -codering, een

aselecte steekproef van 100 teksten getrokken. De respondenten waren allen eerstejaarsstudenten aan wie in de loop van het tweede semester gevraagd is een argumentatieve tekst van 500 woorden te schrijven waarin ze de lezer moesten overtuigen van hun visie op sociale netwerksites. (De opdracht is toegevoegd als bijlage 2). Daarbij is ernaar gestreefd om het kader waarin studenten teksten schrijven maximaal te respecteren. De deelnemers beschikten over één uur om een tekst op pc te schrijven en mochten gebruik maken van alle mogelijke hulpmiddelen.⁴ Er is aan de deelnemende studenten niet meegedeeld dat het een taalonderzoek betrof om een onnatuurlijk grote aandacht voor taal te vermijden. Ook is ervoor gepocht om de schrijfoopdracht niet op voorhand aan de studenten te communiceren. Onzes inziens garandeert deze werkwijze dat alle studenten de tekst in een vergelijkbare situatie geschreven hebben. Bovendien is een onderwerp gekozen dat nauw aansluit bij de interessesfeer van jongvolwassenen. Uit de bevraging na afloop van de schrijfoopdracht blijkt dat de deelnemers de werkwijze niet als problematisch ervaren hebben. Ze waren van mening dat ze voldoende vertrouwd waren met het thema (gemiddelde score van 3,79 en mediaan van 4 op een vijfpuntschaal) en dat de bemeeten tijd volstond (gemiddelde score van 3,61 en mediaan van 4 op een vijfpuntschaal).⁵

Foutencodering

Het opstellen van een empirisch onderbouwde foutenclassificatie veronderstelt een intersubjectieve identificatie en codering van iedere fout in een tekst. Die doelstelling vereist een eenduidige normbepaling en een welomschreven foutconcept.

De normbepaling kan grofweg herleid worden tot een keuze tussen de gebruiksnorm en de gecodificeerde norm. Hoewel de gebruiksnorm, die realistischer en soepeler is

dan zijn gecodificeerde pendant, nauwer aansluit bij het taalgevoel van een taalgemeenschap, worden onderzoekers daarbij voor het probleem van een gebrekkige intersubjectiviteit gesteld. Immers, iedere taalgebruiker hanteert een idiosyncratische normbepaling die niet alleen van de gecodificeerde norm maar ook van de normbepaling van andere taalgebruikers afwijkt. Wij opteren bijgevolg voor de gecodificeerde norm omwille van de intersubjectiviteit en replicerbaarheid. Voor de norm zijn wij bij de volgende naslagwerken te rade gegaan: voor de spelling de *Woordenlijst Nederlandse Taal* (Van Sterkenburg & Beeken, 2005), voor de woordenschat het *Van Dale Groot Woordenboek der Nederlandse Taal* (Boon & Geeraerts, 2008) en voor de grammatika de *Algemene Nederlandse Spraakkunst* (Haeseryn et al., 1997). Iedere inbreuk tegen de regels in die naslagwerken wordt als een taalfout geïdentificeerd.

Systematisch onderzoek naar taalfouten is hoofdzakelijk gericht op het vreemde talenonderwijs. Tono (2003) merkt op dat het definiëren van het concept taalfout geen sinecure is, waardoor er vaak verwezen wordt naar vage noties als het taalgevoel van de moedertaalspreker, hetgeen voor onze doelstelling ongeschikt is. In de trant van het onderzoek naar *learner corpora*, waar authentiek materiaal van (vreemde) taalleerders bestudeerd wordt (Granger, 2003; Pravec, 2002) met als praktisch einddoel het voorstellen van aangepast leer materiaal voor (groepen van) taalleerders (Nesselhaus, 2004), werd aan iedere fout een code toegekend die zowel taalkundige informatie (spelling, hoofdlettergebruik, interpunctie, lexicon, syntaxis, tekstgrammatica, stijl) als informatie over de aard van de fout (redundantie, omissie, foutiefgebruik) omvat.

Op basis van bestaande coderingsschema's (voor vreemde talen) en de praktijkervaring van de onderzoekers is een initiële foutenmatrijs opgesteld. Na een testfase op 20 teksten resulteerde dit in een coderings-

schema met 44 onderscheiden fouten dat als bijlage 1 is toegevoegd.

Onderzoeksvragen en operationalisering

Bij het opstellen van de classificatie van de taalfouten die eerstejaarsstudenten in een professionele bacheloropleiding maken, laten we ons door de volgende onderzoeksvragen leiden:

1. Wat zijn de meest frequente fouten?
2. Wat zijn de meest typische fouten?

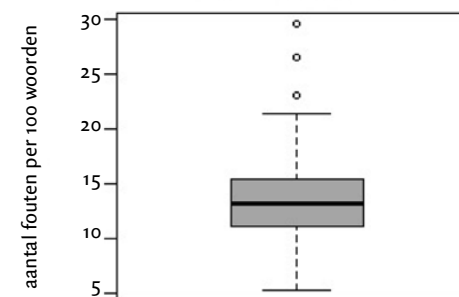
De eerste onderzoeksvraag wordt geoperationaliseerd als de corpusfrequentie (CF) van een fout. Daaronder verstaan we het aantal keren dat een fout in het totale bestand van 100 teksten voorkomt. Op die manier meten we het recurrente karakter van een fout. Door de variabele lengte van de teksten in het corpus is die foutmaat gestandaardiseerd tot het aantal voorkomens van een fout per 100 woorden.⁶

De tweede onderzoeksvraag beantwoorden we aan de hand van de documentfrequentie (DF) van een fout. Dat is het percentage teksten waarin een bepaalde fout voorkomt. Het cijfer geeft een beeld van de verspreiding van de fouten over de teksten, dat wil zeggen de studenten, in de steekproef.

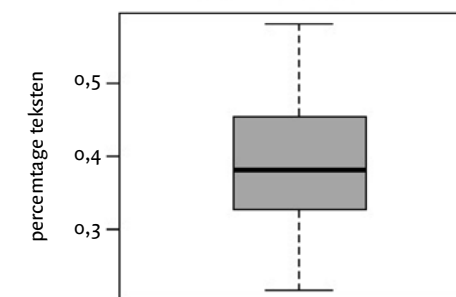
Resultaten

Voordat we de CF- en DF-scores van de verschillende fouten bespreken, werpen we een blik op de verdeling van die foutmaten over de teksten en over de onderscheiden fouten. Dat doen we aan de hand van een boxplot of doosdiagram. Dat is een grafische samenvatting van de centrummaat en de spreiding van gegevens op basis van 5 kengetallen, met name het minimum, het eerste kwartiel, de mediaan, het derde kwartiel en het maximum. Het diagram verdeelt de gegevens in vier gelijke delen en wordt als volgt geïnterpreteerd. De doos (in het grijs) stelt de

VERDELING CF-SCORES OVER DE TEKSTEN



VERDELING DF-SCORES OVER DE TEKSTEN



Figuur 1. Verdeling van de foutmaten over de teksten

centrale 50% van de gegevens voor, waarbij de horizontale lijn in de doos de mediaan identificeert. De verticale lijnen (de zogenaamde snorharen) bovenaan en onderaan de doos geven respectievelijk het hoogste en het laagste kwart van de gegevens weer. Tot slot zijn er de punten buiten de uiteinden van het diagram. Dat zijn uitschieters die extreme waarden identificeren.

De linker boxplot in figuur 1 toont de verdeling van de CF-scores over de teksten. De mediaan voor het aantal fouten per 100 woorden in een tekst bedraagt 13,20. De verdeling vertoont een grote spreiding: terwijl de tekst met de laagste foutenlast gemiddeld 5,32 fouten per 100 woorden telt, treffen we 21,36 fouten aan in de tekst met de hoogste foutenlast, en zelfs 29,49 als we de uitschieters in beschouwing nemen. Die cijfers wijzen enerzijds op een hoge foutenlast. Rekening houdend met een gemiddelde zinslengte van 17,11 woorden moeten we besluiten dat professionele bachelorstudenten moeite hebben om een foutloze zin te schrijven. Anderzijds dient in herinnering te worden gebracht dat de doorsnee taalgebruiker bij het lezen van de teksten een milder oordeel zal vellen aangezien wij de strikte gecodificeerde norm toepassen. De rechter boxplot vat de verde-

ling van de DF-scores over de teksten samen. We stellen vast dat die spreiding varieert van teksten waarin 21,81% van de onderscheiden fouten voorkomt tot teksten waarin men 58,18% van de mogelijke fouten aantreft, met een mediaan van 38,18%. Als we de fouten in functie van de teksten beschouwen, meten we een matige correlatie tussen de CF-score en DF-score ($r = 0,5681$).

Figuur 2 geeft een overzicht van de verspreiding van de CF-score en DF-score over de 44 onderscheiden fouten. De linker boxplot visualiseert de spreiding van de CF-scores over de verschillende onderscheiden fouten. Daarbij stellen we duidelijk vast dat het gros van de onderscheiden fouten laagfrequent is, met maximaal 0,57 voorkomens per 100 woorden, oftewel minder dan 3 voorkomens per tekst van 500 woorden. De helft van de fouten komt niet vaker dan 0,11 keer per 100 woorden voor. Daartegenover staat echter dat er een beperkt aantal hoogfrequente fouten is, meer bepaald 7 uitschieters, die tussen 3 en 6 keer in een tekst van 500 woorden voorkomen. De rechter boxplot geeft een overzicht van de spreiding van de fouten over de teksten. Een meerderheid van de fouten wordt door een minderheid van de studenten gemaakt: de 50% minst verspreide