

Digitale hulp vir akademiese Afrikaans – eerste lesse uit ’n ontwikkelingsprojek



Authors:

Tobie van Dyk¹

Henk Louw¹

Marlies Taljard²

D.J. Cloete¹

Affiliations:

¹School of Languages, Faculty of Humanities, North-West University, Potchefstroom, South Africa

²Department of Afrikaans, Faculty of Social Sciences and Humanities, University of Fort Hare, Mtatha, South Africa

Corresponding author:

Henk Louw,
henk.louw@nwu.ac.za

Dates:

Received: 29 Oct. 2020

Accepted: 21 Apr. 2021

Published: 29 July 2021

How to cite this article:

Van Dyk, T., Louw, H., Taljard, M. & Cloete, D.J., 2021, 'Digitale hulp vir akademiese Afrikaans – eerste lesse uit ’n ontwikkelingsprojek', *Literator* 42(1), a1758.
<https://doi.org/10.4102/lit.v42i1.1758>

Copyright:

© 2021. The Authors.
Licensee: AOSIS. This work is licensed under the Creative Commons Attribution License.

Read online:



Scan this QR code with your smart phone or mobile device to read online.

Digital assistance for academic Afrikaans – first lessons from a development project: This article identifies digital assistance in the development of academic writing skills (specifically in Afrikaans) as a problem that can be partially improved with annotated data from an appropriate corpus. Reference is made to the corpus that ICELDA and SADILAR developed and annotated for informing the development of the software package, Skryfhulp Afrikaans. The emphasis in the article is on the purpose and usefulness of the annotated data within the context of ICELDA and SADILAR's mandates, and the preliminary findings of the first trial run with the data annotation indicating that the project is headed in the right direction, considered from a pedagogical perspective.

Keywords: SaMulCat; Academic Literacy; SADiLaR; ICELDA; corpus; annotation; digital assistance, Skryfhulp Afrikaans; writing pedagogy; error analysis; annotation scheme; undergraduate writing.

Inleiding

Op 13 November 2019 het die Suid-Afrikaanse regering, in ’n verslag oor die afgelope 25 jaar van demokrasie, deurvloesyfers van tersiêre instellings bekendgemaak. Hieruit blyk dit dat slegs 22% van die voorgraadse studentepopulasie ’n driejarige B-graad uiteindelik wel in drie jaar voltooi. Na vier jaar het 39% gegradueer en na vyf jaar 56% (BusinessTech 2019). Uit hierdie data blyk dit ook dat daar nie statisties beduidende skuwe voorgekom het wat betref die gradueringskoerse nie. Daar is vanselfsprekend ’n reeks veranderlikes wat ’n invloed het op die sukseskoers, insluitend die demografiese profiel van die studente, voorafgaande opleiding, sosio-ekonomiese omstandighede, ras, geslag, die tempo en vlak van akademiese akkulturasie en (belangrik vir hierdie artikel) akademiese taalvermoë.

Taal en kommunikasie as deel van suksesvolle akademiese akkulturasie

Kommunikasie in al sy vorme is ’n belangrike menslike aktiwiteit. Die Latynse woord *communicare* beteken ‘om te deel’. Deur kommunikasie deel ons onder meer met mekaar inligting, gedagtes, persepsies en emosies.

Hoewel dit moontlik is om sonder woorde te kommunikeer, is taal die mees algemene vorm van menslike kommunikasie. Die basiese beginsel van Roman Jakobson (Tribus 2017) se bekende kommunikasiemodel veronderstel dat die kommunikasieproses bestaan uit ’n boodskap wat deur ’n sender aan ’n ontvanger gestuur word. Indien die boodskap nie duidelik verwoord en gekonkstualiseer is nie, verstaan die (ideale) ontvanger nie die sender nie. Om in ’n akademiese omgewing suksesvol te kommunikeer, moet in gedagte gehou word dat akademiese diskoers ’n gespesialiseerde vorm van praat, skryf en lees behels. Indien die sender se boodskap nie aan die vereistes van dié diskoers voldoen nie, word dit nie deur die ontvanger verstaan nie, en hou dit ernstige negatiewe gevolge vir die sender in.

Die proses van kulturele en sielkundige verandering in ’n individu deur kontak met ander lede van ’n bepaalde akademiese gemeenskap, word akademiese akkulturasie genoem. Studente, as nuwe en ontwikkelende vakkundiges, kan slegs effektief met die gevinstige gemeenskap kommunikeer wanneer hulle betrokke raak by die diskoers van die gemeenskap. Hyland (2009:1) beskryf dit as ‘ways of thinking and using language which exists in the academy’. Die onvermoë om die toepaslike akademiese diskoers te verstaan en te gebruik, is een van die belangrikste oorsake van akademiese mislukking (Weideman 2003). Akademiese akkulturasie is die vermoë (en motivering) om die akademiese diskoers in al sy diversiteit te assimileer, te verstaan, dit te internaliseer en daarin te kan kommunikeer. Dit is ’n noodsaaklike voorwaarde

vir akademiese integrasie, en dit is dikwels moeilik om by so'n punt uit te kom (Van Dyk 2015). Vir eerstejaarstudente, of jong akademici in die besonder, behels dit om te groei tot 'n volwaardige lid van die betrokke akademiese gemeenskap, deur bewus te word van hoe om die akademiese gemeenskap se kommunikatiewe gesag en oorwig te gebruik (en inderdaad daarby te kan aanpas): norme en praktyke, waardes en verwagtinge, en die taalkundige konvensies wat deel van die bepaalde akademiese diskfers uitmaak (vgl. Duff & Hornberger 2010; Gee 1990; 1998; 2000; Hyland 2009; Van de Poel & Gasiorek 2012).

Wanneer studente die vermoë ontwikkel het om toepaslik in gesprek te tree met lede van die akademiese diskfersgemeenskap (en spesifieke dissiplinêre gemeenskappe binne die groter diskfersgemeenskap), sal hulle waarskynlik beskou word as suksesvolle lede van daardie gemeenskap. 'n Belangrike oorweging in hierdie ontwikkelingsproses is dat die leerders oor die nodige selfvertroue (onder meer gebaseer op hul kennis van en ervaring in die diskfersgemeenskap) moet beskik om gemaklik en gefokus verder te groei.

Wanneer dít realiseer, sou daar aangevoer kon word dat akademiese akkulturasie die sleutel tot akademiese sukses is. Die sukses van akademiese akkulturasie word sigbaar deur die vermoë van studente om hulle sinvol in die skryfmodus uit te druk.

Vanuit 'n sosiolinguistiese perspektief kan akademiese skryfwerk as 'n vreemde diskfers beskou word, oftewel 'n vreemde taal wat aangeleer moet word. Om sinvol deur die skryfmodus te kan kommunikeer, hou onder meer in dat die student ingewikkeld terminologie, wetenskaplike retoriek, kennis van tegniese strukture, metakognisie en kritiese denkvaardighede onder die knie moet hê. Studente moet dus die wetenskaplike norme en konvensies verwerf wat nodig is om in die akademiese gemeenskap te kan oorleef en te funksioneer. Hierdie is van toepassing op beide 'n generiese en dissiplinespesifieke diskfers.

Dit word algemeen aanvaar dat gebrekkige taalvaardigheid (waaronder swak skryfwerk) een van die veranderlikes is wat lei tot die lae deurvloeiwers by universiteite (sien bv. Bizzell 1992; Blanton 1994 herdruk in Zamel & Spack 1998). Daar bestaan trouens geen twyfel dat universiteitstudente op intreevlak ernstige probleme ondervind om hulself skriftelik binne die tersiêre omgewing uit te druk nie. Butler (2007) toon aan dat hierdie onvermoë nie tot voorgaarde studente beperk is nie, maar dat ook nagraadse studente dikwels probleme ondervind wanneer dit kom by die skryf van verhandelings en proefskrifte. Die rede hiervoor is dat taal as voertuig gebruik word om gedagtes te struktureren en te orden tydens die soek, prosessering en produsering van akademiese inligting (Doughty & Pica 1986; Van Dyk 2015). Al drie hierdie prosesse vind plaas deur te luister, lees, praat of skryf in verskillende kombinasies, afhangende van die

behoefte en konteks (Bachman & Palmer 1996; Van Dyk & Van de Poel 2013; Van Dyk & Weideman 2004).

Dit is egter veral tydens die prosessering en produsering van akademiese inligting wat die skryfvaardighede van studente ter sprake kom – hierdie twee prosesse is deurlopend in wisselwerking met mekaar. Om teks te kan produseer (vir die doeleindes van hierdie artikel, verwys dit na skryf), moet inligting ingesamel word, geanalyser word, met ander inligting in verband gebring word en van toepassing gemaak word op ander kontekste. Voorts is die oorgrote meerderheid van akademiese assessorings op skriftelike produksie geskoei (Hyland & Hyland 2006:85). Studente moet deur hierdie modus kan aantoon dat hulle wel die inhoud bemeester het en oor bepaalde vaardighede (kritiese denke, sintese van inhoud, ens.) beskik. Daarom is taal (en skryf in die besonder) van belang hier.

Alhoewel 'n hoë vlak van akademiese skryfvaardigheid van kardinale belang is in akademiese sukses, skryf studente swak, hetsy in hul eerste, tweede of addisionele taal (sien bv. Butler 2007). Boonop is die bemeesterung van inhoud wat in akademiese skryfwerk ten toon gestel moet word, sowel as die vermoë om dit op 'n grammataal korrekte, goed gestructureerde, gepaste, samehangende en vloeiende wyse te produseer (Lasagabaster 2008; Skehan 1988; Spolsky 1973) aansienlik moeiliker as wat van leerders op skool verwag is. Akademiese skryfwerk moet aan ander skryfkonvensies voldoen as waaraan studente tipies gewoond is; verskillende dissiplines vereis boonop soms spesifieke benaderings en praktyke (Hyland 2004; 2008; Swales & Feak 2000).

Ten einde hierdie tekortkominge in studente se monderings te probeer hanteer, stel verskeie universiteite verpligte skryfkursusse in, of bied kursusse in akademiese geletterdheid aan, waarin daar aandag aan akademiese skryfvaardighede gegee word – dit is tipies deel van die *English for Academic Purposes-* en die *Academic Literacies*-bewegings. Die studentegetalle in hierdie kursusse is egter dikwels so hoog dat dit nie altyd moontlik is om individuele aandag aan studente en hul spesifieke skryfprobleme te gee nie.

Beskikbare digitale hulpmiddels (soos spel- en grammatikatoetsers) is dikwels slegs vir Engels beskikbaar, is meestal nie op akademiese skryfstyle ingestel nie, of is nie bekostigbaar vir studente nie. Dit voldoen boonop nie noodwendig aan die individuele skryfbehoeftes van studente nie omdat dit, onder meer, dikwels nie in staat is om plaaslike tweedetaalsprekers se eiesortige foute te identifiseer nie. As dit verder in ag geneem word dat akademiese diskfers, en spesifieker die produksie daarvan, ook geag kan word as 'n addisionele taal, of selfs 'n vreemde taal (Gee 1998; Van Dyk 2010), dan is die probleem van akademiese skryfwerk as sodanig, sowel as die beskikbaarheid van digitale hulpmiddels, selfs nog groter.

Die bestaande blyk 'n internasionale probleem te wees, en verskillende universiteite stel, soos reeds genoem, verskillende ondersteuningsmeganismes aan studente beskikbaar ten einde dit minstens ten dele te probeer oplos. 'n Voorbeeld hiervan is die Katholieke Universiteit Leuven in België (deesdae bekend as KU Leuven) wat 'n projek deur die Instituut voor levende talen (2020) van stapel gestuur het om studente se skryfwerk op akademiese vlak te verbeter, deur van digitale hulpmiddels gebruik te maak. Dié digitale hulpmiddel, getiteld *Schrijfhulp* (De Wachter, D'Hertefelt & Heeren 2016; Meihuizen 2019), is daarop gemik om studente aan dié Nederlandstalige universiteit te help om die algemene foute in hul akademiese skryfwerk te identifiseer en te korrigieer. *Schrijfhulp* is intussen ook uitgebrei na Frans, die ander onderrigtaal aan die KU Leuven, en 'n basiese weergawe is ook in Engels beskikbaar.

Probleem en doel

Daar bestaan tans geen produkte soortgelyk aan *Schrijfhulp* wat voorsiening maak vir die foute wat Suid-Afrikaanse studente in hul akademiese skryfwerk maak nie. Die produkte wat wel bestaan, is (soos ook hier bo gemeld) duur, hoofsaaklik in Engels beskikbaar, en nie noodwendig vir die akademiese konteks ontwikkel nie. Daar is gevvolglik 'n ooreenkoms met die KU Leuven aangegaan om die *Schrijfhulp*-sagteware as deel van 'n loodsprojek aan te pas vir die Suid-Afrikaanse konteks, met spesifieke verwysing na 'n digitale hulpmiddel wat skryfondersteuning in Afrikaans bied. Die produk wat ontwikkel is, heet *Skryfhulp Afrikaans*. Die ooreenkoms is aangegaan tussen die vier vennote van die Inter-Institutional Centre for Language Development and Assessment (2020), oftewel ICELDA, en die KU Leuven.

Vir sodanige sagteware om effektief te kan werk, benodig die program se algoritmes onder meer woordelyste met 'n aanduiding van die lemma, verbuigings daarvan, die woordsoort, moontlike kollokasies, sowel as lyste van foute wat algemeen in studenteskryfwerk voorkom. Alhoewel dosente dikwels 'n goeie aanvoeling het vir die foute wat hul studente maak en dit bloot sou kon lys, is dit ook nodig om 'n akkurate, kwantifiseerbare beeld te vorm van die tipe foute wat studente gereeld maak, ten einde die algoritmes vir *Skryfhulp Afrikaans* se akkuraatheid te verbeter. Om hierdie rede word daar 'n korpus versamel van werklike studentetekste by drie van die vier ICELDA-venootskapsuniversiteite. Dit is 'n korpus waarin die foute sistematies en konsekwent geïdentifiseer kan word (meer inligting volg hier onder). Etiese klaring is by al drie instellings hiervoor verkry.

Hierdie benadering is 'n vorm van foutanalise (*error analysis*), en alhoewel die tegniek algemeen gebruik word in die linguistiek, is daar besonder min gedoen in Afrikaans, veral wat betref akademiese skryfwerk op voorgaardevlak. Die enigste navorsing wat daaroor opgespoor kon word, het te make met die invloed van Afrikaans op Engelse

produksie en foute wat tweedetaalsprekers van Afrikaans maak wanneer hulle Afrikaans produseer (vgl. Roos 1991; Stander 2000). Dit laat 'n gaping in die kennisdomein oor die foute wat Afrikaanse studente maak in hul produksie van geskrewe Afrikaans, veral op akademiese vlak. Voorts is daar nog slegs 'n betreklike klein korpora in Afrikaans beskikbaar, veral indien die vereiste is dat die korpus slegs uit stu-den-te-skryf-werk moet bestaan.

Die probleem, in kort, is naamlik dat daar nie gesikte ondersteuningsagteware vir akademiese Afrikaanse skryfwerk bestaan nie, en dat 'n foutanalise noodsaaklik is om algoritmes in 'n digitale hulpmiddel effektief te laat funksioneer. Die doel van hierdie artikel is daarom om te rapporteer oor 'n ondersoek wat gedoen is oor die tipiese foute wat universiteitstudente in hul akademiese skryfwerk maak, ten einde tot ingeligte pedagogiese besluitneming te kom, wat akademiese skryfondersteuning tot voordeel kan strek, en terselfdertyd *Skryfhulp Afrikaans* se akkuraatheid wat betref die algoritmes, te kan bevestig.

Vooruitskouing

Die grootste deel van hierdie artikel is 'n begronding van die metode vir die projek *Skryfhulp Afrikaans*. Die wyse waarop die korpus saamgestel is, word in besonderhede bespreek in Carstens en Eiselen (2019) en val buite die reikwydte van hierdie artikel. In hierdie artikel word daar wel verwys na ICELDA se mandaat vir taalontwikkeling en taaltoetsing, waarna die doel en nut van geannoteerde data oorsigtelik uitgelig word, om te illustreer hoe dit gebruik kan word om ICELDA se mandaat te ondersteun. 'n Beskrywing van die annotasieprosesse wat in die projek gebruik word (annotasiekema en -sagteware, asook die annoteerders), word aangebied en dit word gevolg deur 'n beperkte bespreking van die resultate en bevindinge vanuit die eerste stel geannoteerde tekste. Die voorlopige bevindinge het uiteraard pedagogiese implikasies, en dui daarop dat daar op mikrovlak baie meer taalfoute voorkom in studente se taalbeheersing as wat tans byvoorbeeld in tipiese akademiese geletterdheidskursusse hanteer word, of wat tipies deur skryfsentra hanteer word in skryfkonsultatsies. Die voorlopige bevindinge kan reeds gebruik word om *Skryfhulp Afrikaans* te verbeter, maar ook om die tegnieke en metodes in ander vorme van skryfondersteuning te verfyn.

Taalontwikkeling binne die SADiLaR-en ICELDA-kontekste

Die South African Centre for Digital Language Resources (2020), oftewel SADiLaR, is een van verskeie sentra binne die South African Research Infrastructure Roadmap (SARIR). Die oorkoepelende doelwit met SARIR (Department of Science and Technology 2016) is:

[T]o provide a strategic, rational, medium to long-term framework for planning, implementing, monitoring and evaluating the

provision of research infrastructures necessary for a competitive and sustainable national system of innovation. (p. 6)

Binne hierdie raamwerk is 13 sogenaamde navorsingsinfrastrukture daargestel. Die meerderheid hiervan fokus op die Natuur- en Gesondheidswetenskappe, sowel as Ingenieurswese, wat in ooreenstemming is met die Nasionale Ontwikkelingsplan 2030. SADiLaR is die enigste infrastruktuur in die geesteswetenskappe en die fokus daarvan is volgens Department of Science and Technology:

[To] acquire [and] systematically develop and manage digital text and speech resources of the official languages of South Africa, creating representative and sustainable resources for R&D. It will also facilitate research capacity building by promoting and supporting the use of digital data, relevant software, and innovative methodological approaches in the humanities and social sciences. (p. 36)

Met hierdie mandaat het SADiLaR 'n naaf-en-speak-model as bestuurstruktuur daargestel. Die naaf, of die kern, van SADiLaR word gehuisves aan die Noordwes-Universiteit se Potchefstroomkampus. Hiervandaan vind alle bestuur plaas, word verskeie aktiwiteite geloods en ondersteuning gebied. Tans is daar vyf nodus wat op verskillende aspekte van SADiLaR se mandaat fokus, naamlik:

1. Die Universiteit van Pretoria se Departement Afrikatale.
2. Die Universiteit van Suid-Afrika se Departement Afrikatale.
3. Die Wetenskaplike en Nywerheidsnavorsingsraad (WNNR) se Human Language Technologies-navorsingsgroep.
4. Die Noordwes-Universiteit se Sentrum vir Tekstegnologie (CTextT).
5. Die Inter-Institutional Centre for Language Development and Assessment (ICELDA).

Die Inter-Institutional Centre for Language Development and Assessment (ICELDA) (<https://icelda.com/>) is die formeke uitkoms van samewerking sedert 2004 tussen vier meertalige universiteite, te wete die Noordwes-Universiteit, die Universiteit van Pretoria, die Universiteit Stellenbosch en die Universiteit van die Vrystaat.

Verskeie ander ooreenkoms tot die stimulering van plaaslike en internasionale samewerking is die afgelope dekade ook gesluit met onder meer die South African Centre for Digital Language Resources (SADiLaR), die Sentrum vir Tekstegnologie (CTextT), die Network of Expertise in Language Assessment (NEXLA), die Council for Quality Assurance in General and Further Education and Training (UMALUSI), die Department of Science and Innovation (DSI), InterCulturate, die Universiteit Antwerpen en die KU Leuven in België, asook die Radboud Universiteit, Nijmegen in Nederland. Voorts is daar verskeie dienste gelewer aan, en gesamentlike projekte aangepak met plaaslike instansies en veral universiteite, ten einde navorsing en produkontwikkeling op die gebied van taalontwikkeling en -assessering te stimuleer.

ICELDA het dit daarom ten doel gestel om onderrig-leerhulpmedele te ontwikkel wat ingeligte pedagogiese besluitneming kan bevorder. Tot op hede het die samewerking tussen die vier vennootskapsuniversiteite in hoofsaak gefokus op die ontwikkeling en implementering van taalhulpbronne vir die taalonderrigsituasie. Dit sluit onder meer die volgende in: taaltoets, onderrig-leermateriaal, ondersteuning in sillabus- en kursusontwerp, ondersteuning met die formulering van taalbeleid en taalplanne, en werkswinkels (byvoorbeeld in akademiese skryfontwikkeling).

Soos hierbo aangedui, is ICELDA (binne die SADiLaR-konteks) een van die nodus. In die ICELDA-nodus word daar in die besonder gefokus op die regverdigte en onsydige taaltoetsing vir spesifieke doeleteindes, en tweedens op die ontwikkeling van intelligente elektroniese hulpbronne vir akademiese skryfwerk. Laasgenoemde is van toepassing op hierdie artikel, met die fokus op *Skryfhulp Afrikaans*.

Die doel en nut van 'n korpus en geannoteerde data

Die gaping wat *Skryfhulp Afrikaans* moet vul, is om studente op pedagogiese wyse te ondersteun met die skryf van akademiese tekste wat grammatikaal korrek, gepas, vlot en samehangend is, met ander woorde om teks te produseer wat aan die idiosinkratiese vereistes van die akademie voldoen. Die pedagogiese vereiste dikteer dat daar ook leiding moet wees en nie net foutaanduiding nie. *Skryfhulp Afrikaans* moet daarom in staat wees om op 'n intelligente wyse foute op verskillende vlakke te kan identifiseer. Dit beteken dat nie net grammatikale foute wat 'n speltoetsers kan uitwys, geïdentifiseer word nie, maar ook foute wat betref taalkennis (ingesluit organisatoriese kennis en pragmatiese kennis), asook strategiese bekwaamheid (metakognitiewe strategieë) – verwys na Bachman en Palmer (1996) vir 'n uiteensetting van wat taalvermoë behels.

Dit was dus nodig om foutlyste op te stel en in die program in te bou as onderbou vir die sagteware om funksioneel te kan wees. Die foutlyste bevat tipiese foute wat universiteitstudente in hul akademiese skryfwerk maak, en dit word sodanig geëenkodeer dat die sagteware dit kan identifiseer en leiding kan gee oor hoe om die taaluiting moontlik anders aan te pak. Spel- en grammatikafoute, sinslengte en foutiewe gebruik van voegwoorde, problematiese makrostruktuur, foutiewe gebruik van kohesie- en koherensiemarkers, en die formaliteit van woordkeuse (styl) is enkele voorbeeld van foute wat deur die sagteware geïdentifiseer kan word. Die akkuraatheid van die foutherkennung hang egter af van die akkuraatheid van die algoritmes, en daarvoor is werklike, kwantifiseerbare data nodig – data wat uit 'n korpus bekom moet word.

Leech (1992:116) dui aan dat 'n korpus nie 'n lukrake samestelling van tekste is nie, maar eerder 'n doelgerigte samestelling van tekste in die poging om 'n

verteenwoordigende beeld van 'n spesifieke groep, taal of tekstile te verkry. McEnery, Xiao en Tono (2006:4) brei hierop uit deur te stel dat 'n korpus aan vier vereistes moet voldoen. Hierdie vereistes is naamlik dat dit masjien-leesbaar moet wees, outentiek, 'n uittreksel (alle instansies van 'n tekstile kan nie ingesluit word nie), en laastens moet dit verteenwoordigend wees. Hierdie korpora kan dan geannoteer word met 'n verskeidenheid linguistiese inligting (Svartvik 1992). Die huidige projek het huis ten doel om annotasies op verskeie vlakke te doen, en met die gepaardgaande metadata kan insiggewende groeperings en analises gedoen word. Met *annotasie* word hier bedoel dat sekere verskynsels in die korpus geïdentifiseer en belig word, of soos McEnery et al. (2006:32) aandui '*undertaking and making explicit a linguistic analysis*'. Dit val buite die bestek van hierdie artikel om 'n volledige oorsig oor die verskillende tipes korpusannotasie te voorsien. Ons volstaan daarby dat die tipe annotasie wat vir die projek aangepak is, pragmatiese annotasie, stilistiese annotasie, probleemgebaseerde annotasie en foutannotasie insluit (McEnery et al. 2006:40–43).

Tot redelik onlangs was dit nog betreklik moeilik om gespesialiseerde leerderdata uit 'n korpus te onttrek, maar huidige tegnologie gee korpuslinguiste toegang tot groot databasisse en kragtige analitiese instrumente. Dit beteken dat statistiese data wat uit so 'n korpus onttrek word, akkuraat en verteenwoordigend behoort te wees. Gilquin et al. (2007:321), sowel as McEnery en Xiao (2011:369–373), toon aan dat data deesdae aan baie meer gevorderde, geautomatiseerde metodes onderwerp kan word. Daardeur word dit moontlik om dit beter te kwantifiseer en te verryk met metadata, asook met 'n wye verskeidenheid linguistiese annotasies, ten einde die korpus se kenmerkende leksikogrammatikale en stilistiese eienskappe te beskryf. Dit is voor die hand liggend dat sodanige inligting waardevolle insigte behoort te lewer – insigte wat die ontwerp van sillabusse, leerdermateriaal en taaltoetse betref.

Verskeie navorsers is dit eens dat korpora belangrike leerinstrumente en linguistiese bronne in taalonderrig is (Flowerdew 2009; Gilquin, Granger & Paquot 2007; McEnery & Xiao 2011; Yoon 2008). Met betrekking tot data wat uit leerderkorpora onttrek is, lyk dit byvoorbeeld asof die intensiewe onderrig van grammaatika in tweedetaal akademiese taalkursusse, 'n noodaakklikheid is. Dit blyk uit statistiese data wat internasionaal uit leerderkorpora onttrek word, dat studente se grammaatikale vaardighede swakker is as wat vantevore algemeen aanvaar is. Voorlopig lyk dit asof dieselfde geld vir voorgaadse Afrikaans eerstetaalstudente (vergelyk die bespreking hieronder). Daar kan geargumenteer word dat die fokus op retoriiese vaardighede in die eerste deel van die sillabusse vir akademiese taalverwerkingskursusse dalk ietwat misplaas kan wees, en eers aandag behoort te geniet wanneer leerders se taalsensitiwiteit, asook hulle leksikale en grammaatikale akkuraatheid opgeskerp is (McEnery & Xiao 2011:367; Yoon 2008:44). McEnery en Xiao (2011:368) beveel byvoorbeeld ook aan dat daar 'n groter klem op die onderrig van

kollokasies geplaas behoort te word. Die rede hiervoor is dat sowel leerders as ervare skrywers dikwels fouteer met die gebruik van idiome, vaste en semi-vaste leksikale uitdrukkings en verskillende grade van buigbaarheid of variansie, sowel as die formaliteitsgraad van dié kollokasies.

Induktiewe korpusgebaseerde leerbenaderings bestaan (sien Flowerdew 2001a:395), maar dit is nie die mikpunt van die korpus of van *Skryfhulp Afrikaans* nie. Die uiteindelike mikpunt is wel die pedagogiese gebruik van die inligting. Soos wat Krishnamurthy en Kosem (2007:358) uitwys, is dit verkiekslik dat studente nie self met rou data of 'n korpus moet werk nie, maar bekwame personeel kan die inligting vanuit 'n korpus gebruik om sillabusse te verryk. Noukeurige korpusanalise behoort dosente in staat te stel om meer praktykgerigte leermateriaal en taaltoetse te ontwerp. Dit word gemotiveer deur besluite wat gegrond is op bevindings afkomstig uit betroubare, gekwantificeerde, empiriese data (Gilquin et al. 2007; McEnery & Xiao 2011). Flowerdew (2001:364) argumenteer byvoorbeeld dat indien 'n leerderkorpus deeglik geannoteer is, dit argumentsonthalwe saam met 'n korpus van akademiese taal gebruik kan word om kurrikuleringsbesluite te neem. Deeglike annotasie van 'n leerderkorpus behoort volgens haar, data te verskaf met betrekking tot tipiese kollokasiepatrone, pragmatiese funksies van teksdele, en diskokerskenmerke in dié korpus. Indien dit met 'n korpus van akademiese taal vergelyk word, behoort dit probleme bloot te lê wat leerders met die skryfproses ervaar. Dit behoort materiaalontwerp, sillabusontwerp en toetsontwerp te onderlê (Charles et al. 2009). Flowerdew (2001) wys daarop dat leerders byvoorbeeld sekere woorde wat in akademiese taal voorkom, ken, maar dat hulle nie bekend is met die leksikogrammatikale patronen (kollokasiepatrone) waarin dit in 'n akademiese diskokers gebruik word nie. Insgelyks kan leerders dalk sukkel om modale hulpwerkwoorde, die passiewe vorm en diskokersmerkers korrek te gebruik, hoewel hulle weet dat dit belangrike aspekte van die akademiese diskokers uitmaak.

Alhoewel daar sommige korpora bestaan wat spesifiek fokus op beginnerskrywers (in verskeie tale), kan dit nie sonder meer aangeneem word dat alle beginnerskrywers dieselfde probleme ondervind nie. Gilquin et al. (2007:322) wys daarop dat sekere skryfprobleme blykbaar algemeen by beginners voorkom soos die oormatige gebruik van die eerstpersoonsvorm. Die verskille, opvoedkundig gesproke, is egter veel belangriker as die ooreenkoms, en daarin lê belangrike metodologiese en opvoedkundige implikasies opgesluit.

Die ontwikkeling van 'n geannoteerde korpus van voorgraadse Afrikaanse akademiese skryfwerk deur ICELDA, behoort daarom belangrike kwalitatiewe en kwantitatiewe data te verskaf aangaande die skryfvaardighede van dié spesifieke groep leerders. Wat die ontwikkeling van 'n geannoteerde korpus van voorgraadse Engels tweedetaalstudente se akademiese skryfwerk deur ICELDA betref, sou in vergelyking

met ander soortgelyke korpora wêreldwyd, interessante lig konwerp op ooreenkomste en verskille ten opsigte van die tipiese taalgebruik van leerders met verskillende moedertale, leerders in verskillende dele van die wêreld, en leerders uit lande met verskillende lingua franca. Die vergelyking van die Afrikaanse korpus met die korpora van ander moedertaalsprekers, sou onder meer, algemene probleme van beginnerskrywers kon uitwys.

Vanuit 'n ander fokus gesien, het geannoteerde korpora ook die voordeel dat dit gebruik kan word in onderwyseropleiding en indiensopleiding (McEnery & Xiao 2011:370). Die annotasie van die Afrikaanse ICELDA-korpus het byvoorbeeld getoon dat sekere foute in studenteskryfwerk waarskynlik toegeskryf kan word aan onderrig van dosente. Dit het byvoorbeeld geblyk dat een spesifieke dosent nie voldoende klem geplaas het op korrekte sinskonstruksies by verwysings nie. Dit het naamlik opgeval dat 'n groot groep studente van hierdie dosent almal dieselfde foute gemaak het met inteksverwysings. Deur dosente insae in 'n geannoteerde korpus te gee, behoort dit duidelik te wees hoeveel foute deur 'n groot aantal leerders gemaak kan word. Onderwyssudente en taaldosente kan uit die korpusanalise insig verkry in hul onderrig-leerbenadering, asook hulle eie skryf- en nasienwerk.

Die stelling sou gemaak kon word dat gespesialiseerde en gelokaliseerde korpora (bv. korpora wat deur spesifieke navorsers met spesifieke doelstellings opgestel word) aan navorsers insig behoort te gee in die linguistiese praktyke van 'n bepaalde gemeenskap of taalgebruikers binne 'n sekere domein. (Ghadessy & Roseberry) Voorgraadse Afrikaanse eerstetaalsprekers wat aan Suid-Afrikaanse meertalige universiteite studeer, is 'n voorbeeld hiervan. So is die korpus dan '*a package of specific linguistic features employed by a group of users*' (Hyland 2009:110). Sodanige korpus kan, aldus Hyland, ook gebruik word om taalvariasie aan te toon. Dit kan byvoorbeeld vergelyk word met nagraadse skryfwerk binne dieselfde domein, akademiese skryfwerk binne ander domeine, ensovoorts.

Verder is dit eweneens gebruiklik om korpora statisties aan te wend deur byvoorbeeld woordelyste (van algemene én vakspesifieke akademiese woorde) en frekwensielyste op te stel. Dit kan dan 'n aanduiding verskaf van die tipiese gebruik en verspreiding van individuele woorde en kollokasies (Coxhead 2000; Gilquin et al. 2007:320–321; Hyland 2009:114; Krishnamurthy 2007:357; Xue & Nation 1984). Moderne sagteware (byvoorbeeld *WordSmith Tools*, *KfNgram*) is in staat om korpora outomatis te ontleed om tipiese gebruikspatrone en -frekwensies aan te toon. Hierdie data is veral nuttig om byvoorbeeld idiomatiese taalgebruik te ontleed, om die oorgebruik en ondergebruik van sekere woorde of frases in studenteskryfwerk te identifiseer, en om die korrektheid van kollokasies na te gaan (Gilquin et al. 2007:321). Hierdie metodes is veral gewild in genre- en registernavoring om die kenmerke van 'n akademiese register binne 'n sekere domein te

beskryf. Dit het ook nuwe belangstelling aangewakker in die studie van teksretorika en pragmatiek (sien bv. Flowerdew & Peacock 2001b). Hyland (2009:126–127) voer aan dat deur data op statistiese wyse te abstraheer en bevindinge te maak vanuit 'n groot korpus tekste, algemeen diskursieve praktyke binne 'n bepaalde domein makliker aangetoon kan word as deur enkele tekste te ondersoek. Hoewel menslike insig in die proses verlore gaan, beperk dié metode egter eweneens subjektiwiteit wat dalk kan voortspruit uit die ondersoek van 'n kleiner korpus tekste (bv. tekste van een klasgroep of een jaargang).

Die bou van 'n korpus vir die SADiLaR-projek

Dus, ten einde die *Schrijfhulp*-sagteware van betroubare data te kan voorsien, moet werklike studenteskryfwerk versamel en annoteer word, en daarom is besluit om 'n korpus te versamel en te annoteer. Die korpus, binne die konteks van die SADiLaR-projek, is nie net op Afrikaans of op voorgraadse skryfwerk geskoei nie. Dit is 'n korpus van akademiese skryfwerk wat meertalig (meer akademiese tale as slegs Afrikaans) van aard is, verskillende genres insluit, teks vanuit verskillende dissiplines insluit, en multivlakkig (vanaf eerstejaar- tot nagraadse vlak) is. Hierdie artikel fokus grootliks op die Afrikaanse komponent van hierdie korpus op eerstejaarsvlak.

Alhoewel daar 'n groot verskeidenheid verskillende analises en annotasies op enige korpus gedoen kan word, is die eerste stel annotasies wat vir hierdie projek vereis is slegs 'n proeflopie, met die doelstelling om 'n geheelbeeld te kry van die mees algemene foute wat in studenteskryfwerk voorkom. Die doel hiervan is dat meer ingeligte besluite ook geneem kan word oor die tipes algoritmes wat in die aanlyn skryfhulpprogram ingebou moet word vir doeleinades van akkurate en bruikbare terugvoer aan studente.

Die annotasie van die SADiLaR-korpus

Annotasieskemas

Die annotasieskema wat vir hierdie projek gebruik is, is afkomstig uit Louw (2006). Louw het reeds 'n studie gedoen oor die mees algemene foute wat in eerstejaarskryfwerk voorkom. Hierdie foutlys is aangevul met addisionele notas wat verkry is vanaf verskeie dosente in die vakgebied akademiese geletterdheid, verbonde aan die NWU, waar daar jaarliks etlike duisende tekste hanteer word.

Vanweë die aard van die verskille in die onderskeie onderrigtale, asook die verskille in die tekste (vlakke en genres), is dit nie moontlik om 'n enkele annotasieskema op te stel wat vir elke voorsienbare fout in 'n teks voorsiening maak nie. Boonop gebeur dit dikwels dat dieselfde fout op meer as een manier geïnterpreteer en gekorrigeer kan word. Hiermee saam is daar nuanses van foute wat voorkom. 'n Dubbelsinnige sin kan byvoorbeeld dubbelsinnig wees as gevolg van 'n homofoonfout,

formuleringsfout, swak leestekengebruik, verkeerde sinskonstruksie (wat op sigself verskeie oorsake het), woordverwarring, of 'n kombinasie van bogenoemde oorsake. 'n Meer komplekse tipe fout is byvoorbeeld wanneer 'n inleiding vir 'n akademiese teks foutloos is wat sinskonstruksie en woordkeuse betref, maar tekort skiet wat inhoud betref. Dit bevat dalk geen vooruitskouing nie. Die ideaal is daarom om 'n optimale stel annotasies te skep wat in die meeste gevalle bruikbaar is.

'n Tweede beperking van die annotasieskema, is dat Louw (2006) se foutlys op Engelse tekste gebaseer is. Aanvullings was daarom nodig om vir foute voorsiening te maak wat eie aan Afrikaans is. Hoewel dit een korpus van akademiese skryfwerk is, is tekste in beide Afrikaans en Engels beskikbaar in hierdie korpus. Beide tale word aan die hand van 'n enkele annotasieskema geannoteer. Daar kan gevolglik ook vergelykings tussen Engelse en Afrikaanse foutpatrone gemaak word. Binne die bestaande annotasieskema is daar 'n aantal kategorieë wat nooit gebruik sal word wanneer Engelse tekste geannoteer word nie, omdat dit slegs op Afrikaans van toepassing is, en vice versa.

Annotationsagteware

Data-annotasie is 'n baie komplekse en arbeidsintensieve aktiwiteit, en die annoteerders vir hierdie projek is nie rekenaarlinguiste nie, maar taalkundiges. Daarom is daar in hierdie projek gebruik gemaak van *MarkWrite* as annotasieprogram, alhoewel dit nie primêr ontwerp is vir annotasies nie. Vir die doeleindes van die projek waaroer daar hier gerapporteer word, is dit wel geag as meer gebruikersvriendelik as die bestaande toegewyde annotasieprogramme, wat ook nie noodwendig vir Afrikaanstalige kwessies voorsiening maak nie. *MarkWrite* (sien Louw 2011) is in die eerste plek ontwikkel met die doel om semi-gestandaardiseerde terugvoer aan studente te verskaf. Terselfdertyd word data-annotasie op die tekste gedoen om 'n plaaslike (of persoonlike) korpus te kan bou. *MarkWrite* voeg outomatis die XML-kodes in wanneer 'n annoteerder 'n spesifieke fout identifiseer; sodoende kan die annoteerders werk met die foutlys se beskrywende kategorieë in plaas daarvan om met die rou XML-kodering gemoeid te wees. Die foutlyste wat in *MarkWrite* gebruik is, het reeds voorsiening gemaak vir XML-annotasie.

TABEL 1: Struktuur van die annotasieskema.

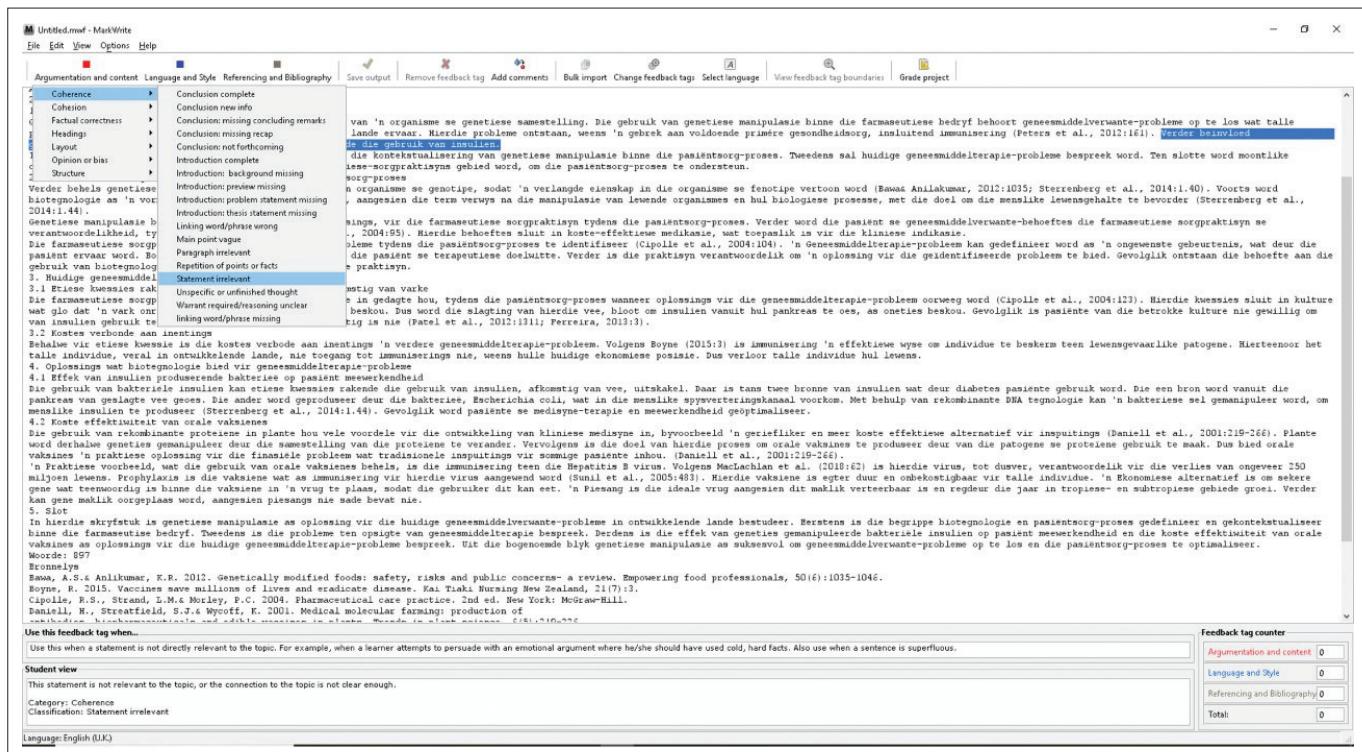
Superordinate	Subordinate	Error type	Feedback	Explanation	XML_label
Language and Style	Argumentation	Paraphrase	This quote is not so unique or difficult that you are unable to paraphrase it. It would be better if you paraphrase it.	Use this when a learner inserted a direct quote instead of paraphrasing, when a paraphrase would have been possible and more desirable.	LSAP
Referencing and Bibliography	Bibliography	Incorrect / Incomplete	This is not the correct way to write a bibliographical entry for this type of source. Consult your guide or ask for assistance.	Use this for an incorrect bibliographical entry. You may add a note to indicate smaller transgressions.	RBBI
Language and Style	Capitalisation	Capitalisation	Your use of capital letters or lower case letters is incorrect.	Use this when learners randomly or incorrectly use Capitalisation. Examples: 'There are plenty of Banks in my country'. or 'I am going to visit my friend in the usa'.	LSCC

MarkWrite se foutlys volg 'n linguisties verantwoordbare struktuur. Vir elke fout of kommentaar in die foutlys, is daar 'n superordinaat (hoofklassifikasie), domein (meer spesifieke klassifikasie), naam vir die spesifieke fout, beskrywing van die fout, unieke XML-kode, verduidelikende nota vir die annoteerder ('Gebruik hierdie merker wanneer ...') en 'n terugvoernota na studente. Die terugvoernota is irrelevant in die konteks van hierdie artikel, maar dit kan wel funksioneer as addisionele verduidelikende nota vir die annoteerders. Daar is drie superordinaatvlakke: argumentasie en inhoud, taal en styl, verwysings en bibliografie. Op die vlak van domein is daar 16 klassifikasies: argumentasie, bibliografie, hooflettergebruik, kohesie, koherensie, feitlike korrektheid, grammatika, opskrifte, taal en styl, uitleg, opinie en vooroordeel, leestekengebruik, verwysings, struktuurfoute, sintaksisfoute en woordvlakfoute. In totaal is daar gewerk met 94 tipes foute. Aangesien dit egter die eerste fase van annotasie was, is die foutlys intussen verfyn en uitgebrei na aanleiding van kommentaar van die gebruikers, sowel as bevindinge van werklike gebruiksfrekvensie. Die mikpunt is om 'n foutlys daar te stel wat nie oormatig baie kategorieë het nie, maar wel die gebruiker in staat sal stel om die optimale hoeveelheid foute akkuraat te kan annoteer.

Tabel 1 (*Struktuur van die annotasieskema*) wat volg, is 'n uittreksel uit die annotasielys wat vir hierdie projek gebruik is. Dit toon die struktuur van drie spesifieke foute in die skema aan, naamlik die onoordeelkundige gebruik van 'n *verbatim*-aanhaling, 'n verkeerde of onvolledige bronverwysing en 'n hoofletterfout.

Binne die konteks van die *Corpus Encoding Standard* ('n stel riglyne vir die optimale annotasie van korpora vir gebruik in taaltegnologiese toepassings), val hierdie tipe annotasies op die vlak van '*mark-up for sub-paragraph structures*' (McEnery et al. 2006:24), alhoewel daar ook oorvleuelings is met '*gross structural mark-up*', aangesien daar na die teks as geheel ook gekyk word vir doeleindes van koherensie. Die annotasieskema word in 'n spreitabel opgestel (soos gewys in Tabel 1), maar dit vertoon binne *MarkWrite* as 'n gewone keuselys soos dié waarmee gebruikers van *Microsoft Office* vertroud is.

Die annoteerder moet in *MarkWrite* die stukkie teks waarin die fout voorkom, selekteer en dan eenvoudig na die korrekte keuse in die keuselys navigeer vir die fout om geannoteer te word (sien Figuur 1: '*MarkWrite* Keuselys' wat volg).



FIGUUR 1: MarkWrite Keuselys.

Wanneer die annooteerder die teks volledig deurgewerk het, word die spesifieke teks in drie formate gestoor, te wete die skoon teks, die teks met XML-annotasies wat vir rekenaars deursoekbaar is (sien Figuur 2 – *XML-uitset van Afrikaanse teks nommer 113*) en ook as HTML-dokument waarin die foute duidelik sigbaar is vir menslike leser.

Empiriese ondersoek

Vervolgens word oor beide Afrikaans en Engels gerapporteer, omdat akademiese skryfwerk op voorgraadsevlak as sodanig ondersoek word. Die vraag is dus nie of dit 'n taalspesifieke skryfprobleem is nie, maar eerder of dit 'n probleem van skryfontwikkeling en denkprosesse is wat daarvan verband hou.

Die annotasies vir die eerste proeflopie is gedoen deur vier annooteerders, twee vir Afrikaans en twee vir Engels. Al vier hierdie annooteerders het opleiding ondergaan in die gebruik van die annotasiekema en -sisteem. Vir beide Afrikaans en Engels is 100 tekste ewekansig geselekteer uit die tekste wat teen die einde van 2018 in die korpus beskikbaar was. Dieselfde 100 tekste in Afrikaans is toe onafhanklik deur twee Afrikaanse taalkundiges geannoteer en dieselfde 100 Engelse tekste is insgelyks onafhanklik deur twee Engelse annooteerders geannoteer. Vir beide Afrikaans en Engels is die annooteerders moedertaalsprekers. Die een Afrikaanse annooteerder is 'n taalkundige met 'n MA-kwalifikasie en uitgebreide ervaring in skryflaboratoriumwerk met voorgraadse studente. Die ander annooteerder is 'n ervare letterkundige met 'n PhD-kwalifikasie, 'n Honneurs in Taalpraktik en sowat 40 jaar

ervaring in skool- en tersiêre opvoeding. Die twee Engelse annooteerders is 'n uitgetrede professor in letterkunde en 'n professionele teksredigeerder met haar eie praktyk. Die doel hiermee was om 'n lys van die mees algemene skryffoute te vind sodat ingeligte pedagogiese besluitneming kan plaasvind.

MarkWrite gee vir elke teks 'n opsomming van die annotasies. Hierdie opsomming is die 'Feedback tag output' en 'Feedback tag counter' (Figuur 3 – Foutlys vir Afrikaanse teks 6 in *MarkWrite-uitset*), asook die statistiek per teks (Figuur 4 – Statistiek vir Afrikaanse teks 6 in *MarkWrite-uitset*). Beide hierdie figure sluit inligting in van een teks as voorbeeld ter illustrasie. Alle geannoteerde tekste se data is uiteindelik ook in 'n spreitabel versamel, ten einde verdere statistiese analises te vergemaklik.

Data

Die eerste 100 tekste in beide Afrikaans en Engels (rofweg 50 000 woorde per taal) het 'n datastel gelewer met 19 231 annotasies. Die getalle per domein word in Tabel 2 (*Annotasiegetalle per domein in die eerste 100 Afrikaanse en Engelse tekste*) weergegee.

Weens die rykheid en omvang van die data, is 'n diepgaande bespreking en analise daarvan nie moontlik binne die beperkinge van 'n enkele artikel nie. Vervolgens enkele voorbeeldteks ter illustrasie: Die domein 'Sintaksis' het byvoorbeeld 16 onderafdelings (wat onder meerwoordorde, onvoltooide sin, aaneenlopende sin, en dubbelsinnigheid insluit). Elkeen van hierdie is ondersoekswaardig. So byvoorbeeld toon 'n nadere analise van die domein 'Sintaksis'

```

TextOnly.113.output - Notepad
File Edit Format View Help
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
<markup>
    <text>
        <tag main="Argumentation and content" tag ="Heading incorrect" code ="AHHI" colour="#FF3333" project="__454__"/>
        **[Is ek vir of teen
genetiese manipulasie?]***
    </text>
    <text>
        <tag main="Argumentation and content" tag ="Good heading" code ="AHHG" colour="#FF3333" project="__454__"/>
        **[Inleiding]**
    </text>
    <text>
        <tag main="Language and Style" tag ="Word order wrong" code ="LSSW" colour="#0066FF" project="__454__"/>
        **[Genetiese
manipulasie word ]**gesien as &apos;n etiese kwessie in die samelewings vandag.
Genetiese manipulasie word gesien as die proses waar genetiese materiaal by &apos;n
organisme bygevoeg word sodat afwykings van gene voorkom kan word. Alhoewel die
proses gebruik maak groot potensiaal te ontwikkel op gesondheid en natuur te
bevorder, is daar teenstanders wat glo dat die mensdom op hierdie manier die
voorwerp of selfs die organisme beheer en manipuleer kan word soos hulle dit
verkies tot hulle eie voordeel. Een van die baie redes maar ook die hoof rede
vir die gebruik van genetiese manipulasie is vir die verbeterde gesondheid en
oorlewing van sekere plante en diere. Die volgende redes van hoekom genetiese
manipulasie voordele vir die mensdom inhoud, gaan verder bespreek word in
diepste. Dit toon groot belofte vir die behandeling van siektes. Dit kan groter
hoeveelheid produksie van voedsel lever deur gene te manipuleer om vrugte en
groente se uitsette te verdubbel. Kloning vorm deel van die belangrikheid van
manipulasie omdat dit aan ons meer inligting verskaf oor ander verwante
tegnologieë.
    </text>
    <text>
        <tag main="Argumentation and content" tag ="Warrant required/reasoning unclear" code ="ACWR" colour="#804D99" project="__454__"/>
        **[Genetiese
manipulasie word ]**gesien as &apos;n etiese kwessie in die samelewings vandag.

```

FIGUUR 2: XML-uitset van Afrikaanse teks nommer 113.

Feedback tag output			
Category	Classification	Feedback	Comments
Syntax	Sentence incomplete	This is not a complete sentence.	
Punctuation	Punctuation missing	You need to insert the appropriate punctuation mark here.	
Language style	Informality	The register in this part of the text is not consistent with the rest of your text or is unsuitable for the context. Beware of being overly formal or informal.	
Syntax	Word order wrong	Sentence word order wrong	
Punctuation	Punctuation missing	You need to insert the appropriate punctuation mark here.	
Syntax	Sentence vague	You leave out too much information for this sentence to make sense. Be more detailed.	
grammar	Auxiliary verb missing	You need an auxiliary verb here	
Factual correctness	Generalisation/unsupported statement	The statement made does not contain relevant facts or explanations to support it. Elaborate further.	
Coherence	Repetition of points or facts	Consider if this is (A) relevant here or if it (B) adds clarification (rephrasing to add strength to your argument).	
Coherence	Statement irrelevant	This statement is not relevant to the topic, or the connection to the topic is not clear enough.	
Opinion or bias	Opinion/personal involvement	This part of your writing shows personal involvement or bias. Academic writing normally requires you to stay uninvolved with the text and/or to be unbiased. Do not use opinions as facts.	
Factual correctness	Generalisation/unsupported statement	The statement made does not contain relevant facts or explanations to support it. Elaborate further.	
Factual correctness	Generalisation/unsupported statement	The statement made does not contain relevant facts or explanations to support it. Elaborate further.	
Layout	Layout wrong for text type	Your layout is not right for this specific text type.	
Structure	Linking sentence missing	You need a linking sentence here to move from the prior paragraph to this one.	
Structure	Linking sentence missing	You need a linking sentence here to move from the prior paragraph to this one.	
Word	Word choice obscuring meaning	This word is not clear enough. Find a better word to say what you want to say.	
Spelling	Combine word	This should be spelled as one word.	
Factual correctness	Generalisation/unsupported statement	The statement made does not contain relevant facts or explanations to support it. Elaborate further.	
Referencing	Reference omitted	You should indicate correctly where you obtained this information by referencing it. Beware of plagiarism.	
Referencing	Reference omitted	You should indicate correctly where you obtained this information by referencing it. Beware of plagiarism.	
Bibliography	Incorrect / Incomplete	This is not the correct way to write a bibliographical entry for this type of source. Consult your guide or ask for assistance.	
Feedback tag counter			
Argumentation and content	10		
Referencing and Bibliography	3		
Language and Style	9		

FIGUUR 3: Foutlys vir Afrikaanse teks 6 in *MarkWrite*-uitset.

dat woordordefoute 220 keer in Afrikaans uitgewys is en 684 keer in Engels, terwyl oorbodige woorde in Afrikaans 152 keer en in Engels 359 keer uitgewys is. Die gebruik van oorbodige woorde in sinskonstruksies hou moontlik verband

met die pasgenoemde onnoukeurigheid in woordkeuse. Net so kan die probleme met woordorde dalk te make hê met die feit dat die meeste studente wat in Engels skryf, Engels as 'n tweede taal 2 gebruik, terwyl die Afrikaanse studente meer



FIGUUR 4: Statistiek vir Afrikaanse teks 6 in *MarkWrite*-uitset.

TABEL 2: Annotasiegetalle per domein in die eerste 100 Afrikaanse en Engelse tekste.

Foutkategorie: Domein	Afrikaans	Engels
Sintaksis	1 787	1 560
Koherensie	1 626	1 150
Woordvlakfout	1 496	1 747
Verwysings	1 112	433
Leestekengebruik	883	2 998
Taal en styl	724	465
Spelling	483	357
Feitelike korrektheid	475	34
Opskrifte	432	12
Kohesie	399	52
Struktuur	234	2
Bibliografie	208	89
Argumentasie en inhoud	206	197
Verwysings	199	197
Hooflettergebruik	102	273
Grammatikafoute	89	408
Vertalingsfout	62	66
Opinie en vooroordeel	14	21
Totale annotasies per taal:	9 894	9 337
Groottotaal gekombineerd:	-	19 231

dikwels in hul eerste taal skryf. Dit sal egter met 'n wetenskaplike ondersoek bevestig of weerlê moet word.

Verder kan daar in 'n enkele sin ook etlike foute voorkom, wat dan alles in *MarkWrite* geannoteer kan word (dit word *klusterfoute* genoem). Voorbeeldel sluit in 'n sintaksisfout, leestekenfout, en selfs 'n kohesie- of grammatikafout. Voorts moet die metadata ook by die interpretasie van die data in ag geneem word. So byvoorbeeld is dit moontlik dat die Afrikaanse tekste van eerste-huistaalsprekers afkomstig is en die Engelse tekste hoofsaaklik van addisionele-taal-sprekers. Dit is daarom interessant dat die Afrikaanse tekste meer sintaksis- en koherensie-foute bevat as die Engelse tekste, en dit verdien 'n verdere ondersoek in fyner besonderhede.

Dit is onomwondne duidelik dat studente se vermoëns te kort skiet met ten minste twee kardinale kennisvlakke in

terme van akademiese taalgebruik: om korrekte, verstaanbare sinne te skryf, en om gebruik te maak van 'n korrekte, akkurate woordeskot. Byvoorbeeld het woordkeusefout ('n subkategorie van die domein 'Woordvlakfout') meer as 700 keer voorgekom in beide die Afrikaanse en Engelse tekste, terwyl die verkeerde woordvorm 178 keer voorgekom het in Afrikaans en slegs 9 keer in Engels. Vermoedelik is Engelse grammatiskatoetsers meer akkuraat om verkeerde woordvorme te korrigéer as die Afrikaanse weergawes daarvan (onder meer weens die komplekse aard van Afrikaanse samstellings). Woordkeusefout het te maken met verkeerde woorde, vae en onduidelike woordgebruik en onakkurate woordgebruik – alles belangrik vir korrekte akademiese kommunikasie.

Pedagogiese implikasies

Die voorafgaande data hou pedagogiese implikasies in, onder meer vir 'n akademiese geletterdheidskursus, skryfsentra en akademici buite die toegepaste linguistiek. Soos reeds genoem, word suksesvolle gebruik van korpora in akademiese taalverwerwingskursusse reeds gerapporteer (sien bv. Flowerdew 2009; Hyland 2009; McEnery & Xiao 2011).

Louw (2006) se lys van die mees algemene foute van eerstejaarstudente en Meihuizen (2019:204) se lys van frekwente skryfprobleme wat deur dosente in 'n akademiese geletterdheidskursus saamgestel is, is voldoende vir kursusontwikkeling en die opstel van nasienskemas. In die konteks van hierdie artikel is die doel egter juis om 'n annotasiekema saam te stel, wat verder toegelig kan word deur foute wat empiries geïdentifiseer word in 'n korpus. Tydens die annotasieproses is die annotasiekema verfyn deur die annoteerders om voorsiening te maak vir foute wat op 'n gereelder basis na vore kom in hierdie studentetekste. Die annotasiekema kan gebruik word om die uitkomstes in 'n akademiese geletterdheidskursus beter te bepaal, na aanleiding van die data wat verkry word uit die annotasies.

Carstens en Eiselen (2019) lewer 'n diepgaande bespreking oor die metadata wat versamel is vir hierdie korpus. Vir die doeleindes van kursusse in akademiese geletterdheid, sal navorsers verder ondersoek kan instel na foute wat voorkom onder verskeie groepe in die korpus: huistaal- of tweedetaalspreker, opvoedkundige agtergrond, geslag, en nog meer. Op hierdie wyse kan kursusse fyner ontwerp word om bepaalde studente beter te ondersteun. Uiteindelik sal 'n individu-gefokusde benadering geneem kan word, indien die kursus 'n spesifieke teikengroep het. Sodoende kan meer effektiewe en pasgemaakte onderrig vir die student gegee word. Hierdie benadering sluit ook aan by ICELDA wat tot op hede in hoofsaak gefokus het op die ontwikkeling en implementering van taalhulpbronne vir die taalonderrigsituasie. Alhoewel dit verskeie aspekte insluit, word daar hier klem geplaas op ondersteuning in sillabus- en kursusontwerp. Vergelyk ook hier Huang (2017:384) rakende die gebruik van korpora vir pedagogiese doeleindes en die afdeling, *Die doel en nut van 'n korpus en geannoteerde data*, wat ook hierop uitgebrei het.

Buiten kursusse wat fyner ontwerp kan word, sal dosente ook op 'n meer effektiewe wyse terugvoer vir studente kan gee. Indien daar byvoorbeeld gekyk word na die tipe skryfbegeleiding wat *Skryfhulp Afrikaans* reeds vir studente kan bied, sal die uiteindelike terugvoer wat daar van 'n dosent verwag word, anders lyk.

Volgens De Wachter et al. (2016:49) is *Schrijfhulp* ontwikkel uit die behoefte dat studente skryfbegeleiding moet ontvang wat hulle tydens die skryfproses ondersteun, maar ook om dosente die geleenthed te bied om meer effektiewe terugvoer te kan gee. Met ander woorde, *Schrijfhulp* identifiseer foute in 'n student se teks wat dikwels voorkom, waarna die student die regstelling kan maak sodat die dosent nie telkens hierdie 'oppervlakkige' (Louw 2007) probleme aan die studente hoef uit te wys nie, maar eerder op die moeiliker inhoudelike aspekte kan kommentaar lewer.

Skryfsentra kan ook baat by die data wat ingewin word deur die annotering van hierdie studentetekskorpus.

Opleidingsmateriaal wat beskikbaar gestel word vir skryfkonsultante, kan meer gefokusseer word sodat skryfkonsultante bepaalde foute makliker kan identifiseer en meer effektiewe terugvoer aan die studente kan gee. Dit is ook makliker om egte voorbeelde in die korpus op te spoor om tydens opleiding te gebruik. Daar is 'n duidelike wisselwerking tussen die opleiding van die konsultant en die ondersteuningshulpbronne wat konsultante tydens 'n skryfkonsultasie gebruik. Dit is daarom belangrik dat hierdie konsultante ingelig word deur middel van opleidingsgeleenthede binne die skryfsentrum, oor die skryfprobleme wat vervat is in die annotasiekema, en uiteindelik die data wat daardeur gegeneer is.

Met die regte opleiding sal die skryfkonsultant tydens 'n konsultasie die gepaste ondersteuningshulpbronne vir die

student kan gee, wat onder meer uitdeelstukke, videoskakels en aktiwiteite insluit. Hierdie ondersteuningshulpbronne kan, na verdere verwerking van die basiese uiteensetting van die data hierbo, geskep word deur byvoorbeeld studente te laat let op woordkeuse en te onderrig in strategieë om meer presies te skryf.

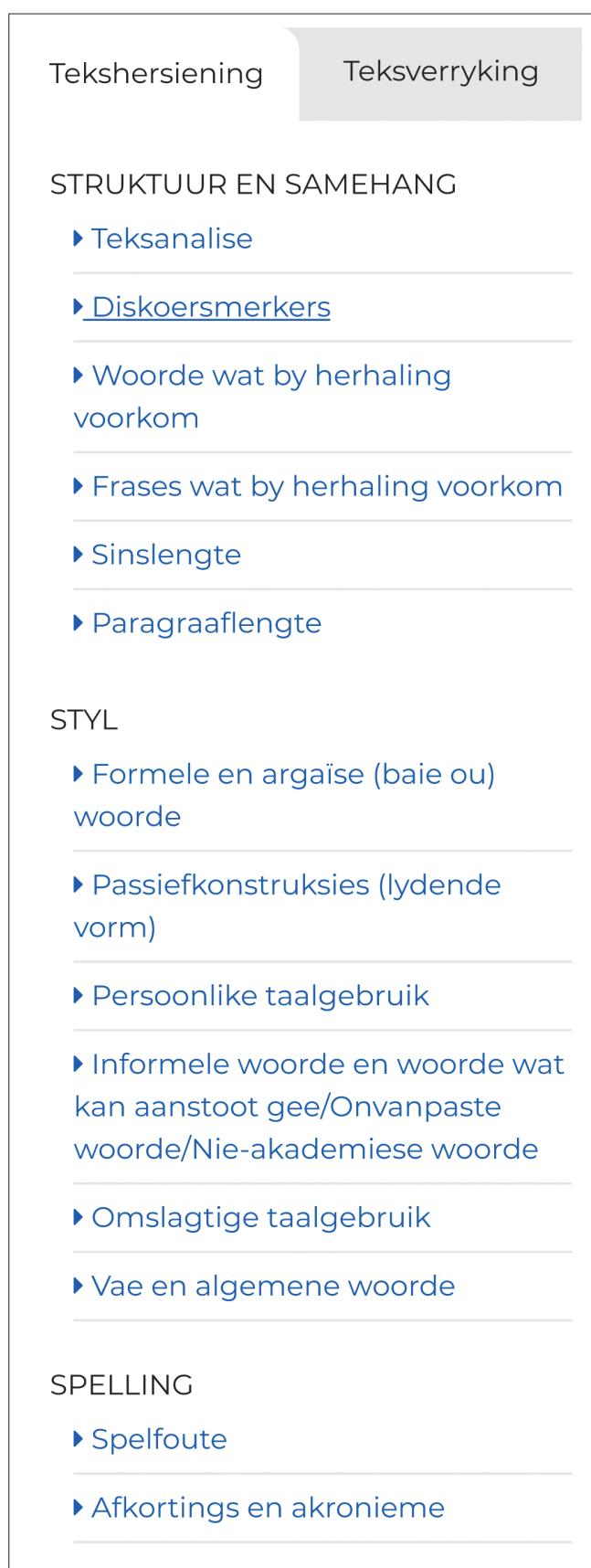
Daar is byvoorbeeld in 2020 reeds 'n beduidende hoeveelheid (9x) addisionele woordeskatalogontwikkelingsoefeninge ingebou in die NWU se aanbieding van akademiese geletterdheid, as wat in 2019 die geval was.

Buite die akademiese geletterdheid-en skryf-sen-trum-omgewing, kan hierdie geannoteerde tekste nuttig wees vir ander kursusse, omdat dit ook tekste insluit van verskillende genres en uit verskeie dissiplines (sien afdeling: *Die bou van 'n korpus vir die SADiLaR-projek*). Met ander woorde, daar word tekste van alle veldes by hierdie korpus ingesluit en tesame met die metadata waartoe navorsers sal toegang hê, kan 'n kursus in enige veld hierby baat, hetsy dit ingenieurswese, kommunikasiestudie of ekonomiese wetenskappe is. Ekonomie-dosente sal vanuit die data tipiese taalfoute kan identifiseer wat hulle studente maak. Ondersteuningsmateriaal of inhoud wat deel vorm van die kursus kan dan ontwikkel word wat pasgemaak is vir hulle omgewing. Ahangende van die onderwerpe, kan ander fakulteite se navorsers ook die inhoud van tekste bestudeer – so stel die Fakulteit Opvoedkunde reeds belang om studente se opinies oor selfgerigte leer te ondersoek, nadat dit aan die lig gekom het dat etlike van die tekste in die korpus oor hierdie onderwerp handel.

Die data is boonop toepaslik op die verbetering en verfyning van die *Skryfhulp Afrikaans*. Alhoewel daar al 'n groot deel van dié 'oppervlakkige' foute deur die skryfhulpmiddel geïdentifiseer kan word, is daar soos met alle tegnologiese produktes, die moontlikheid vir verbetering. Vanuit die data sal daar bepaal kan word welke taalfoute nog nie ingesluit is nie, maar ook hoe die program die student se skryfproses beter kan ondersteun.

Skryfhulp Afrikaans bied tans vir die student terugvoer rondom twee aspekte: tekshersiening en teksverryking (sien Figure 6 – *Keuselys vir die tekshersieningsafdeling in 'Skryfhulp Afrikaans'* en 7 – *Keuselys vir die teksverrykingsafdeling in 'Skryfhulp Afrikaans'*). Wanneer die student op die tekshersiening-oortjie klik, kom die volgende opsies vir moontlike terugvoer aan bod: struktuur en samehang, asook styl, en spelling. Elkeen van hierdie drie kategorieë word onderverdeel in verdere kategorieë. Sien die skermgreep (Figuur 6 – *Keuselys vir die tekshersieningsafdeling in 'Skryfhulp Afrikaans'*) vir 'n volledige uitleg.

In Tabel 2 (*Annotasiegetalle per domein in die eerste 100 Afrikaanse en Engelse tekste*) kan gesien word dat koherensie en kohesie onderskeidelik 1626 en 399 keer in die Afrikaanse tekste gemerk is. Uit hierdie keuselys (Figuur 6 – *Keuselys vir die tekshersieningsafdeling in 'Skryfhulp Afrikaans'*) kan

**FIGUUR 6:** Keuselys vir die tekshersiensafdeling in *Skryfhulp Afrikaans*.

gesien word dat *Skryfhulp Afrikaans* reeds fokus op aspekte soos diskopersmerkers, wat belangrik is om kohesie en koherensie binne 'n teks te bewerkstellig. Verder dui

**FIGUUR 7:** Keuselys vir die teksverrykingsafdeling in *Skryfhulp Afrikaans*.

Skryfhulp Afrikaans ook woordherhaling en sinslengte aan wat te make het met woordkeuse, en uit die tabel kan gesien word dat woordvlakfoute 1 496 keer geannoteer is. Hierdie woordvlakfoute, of ook woordkeuseprobleme, kan verder verbeter word deur die teksverrykingsafdeling van *Skryfhulp Afrikaans*. Dit is die tweede oortjie waarop studente kan klik wanneer hulle van hierdie skryfhulpmiddel gebruik maak. Sien die skermgreep (Figuur 7 – Keuselys vir die teksverrykingsafdeling in 'Skryfhulp Afrikaans') vir die keuselys.

Hieruit kan gesien word dat *Skryfhulp Afrikaans* reeds die probleme wat die geannoteerde data uitwys, kan hanteer. Probleme wat as oppervlakkig beskou word, kan reeds deur so 'n hulpmiddel onderskep word.

Daar kan nie net bloot aanvaar word dat 'n student die nodige woordeskat en kennis het om goeie akademiese sinne te formuleer nie. 'n Akademiese geletterdheidskursus, tesame met ander ondersteuning soos die voorgenoemde *Skryfhulp Afrikaans* kan dien as die vertrekpunt om 'n student verder te ondersteun in die akademiese omgewing. Dit blyk wel uit die data in Tabel 2 (*Annotasiegetalle per domein in die eerste 100 Afrikaanse en Engelse tekste*) dat daar, benewens 'n kursus in akademiese geletterdheid, ook in sommige gevalle gronde is vir 'n addisionele taalverwerwingskursus om die basiese taalkundige onderbou van sinskonstruksie en woordeskat by studente te vestig.

Slot

Die artikel se vertrekpunt is dat dit van kardinale belang is vir akademiese sukses om studente se skryfvaardighede te

verbeter. Dit is egter 'n haas onbegonne taak weens die groot studentgetalle wat verpligte taalontwikkelingskursusse neem. Om hierdie rede moet digitale hulpmiddelle gebruik word en oordeelkundige besluite geneem kan word rakende pedagogie. Die digitale hulpmiddels wat beskikbaar is, beantwoord egter nie aan akademiese vereistes nie en daarom moet nuwe toepassings geskep word – nuwe toepassings wat op werklike en kwantifiseerbare data gegrond is. *Skryfhulp Afrikaans* is so 'n toepassing. Uit die baie beperkte eerste stel geannoteerde data is dit reeds duidelik dat *Skryfhulp Afrikaans* op die regte fondasie gebou is, met die program se klem veral op woordeskat en sintaksis. Aanpassings in terme van pedagogie kan ook reeds gemaak word, en is in sommige gevalle reeds gemaak. Verdere analises van meer data en fynere analises van die data sal tot meer insigte lei. Dit kan ongetwyfeld positiewe effekte in die onderrig van akademiese Afrikaans tot gevolg hê. Verdere navorsingsmoontlikhede sluit in ondersoeke na inter-merker betroubaarheid en die verfyning van die annotasieskema – dit behoort by te dra tot die optimalisering van 'n hulpmiddel soos *Skryfhulp Afrikaans*.

Erkenning

Mededingende belang

Die outeur verklaar dat daar geen finansiële of persoonlike verbintenis is met enige party wat hulle nadelig kon beïnvloed in die skryf van hierdie artikel nie.

Outersbydrae

Alle skrywers het ewe veel tot hierdie werk bygedra.

Etiese oorweging

Hierdie artikel volg alle etiese standarde vir navorsing.

Befondsing

Hierdie navorsing het geen spesifieke toekenning ontvang van enige befondsingsagentskap in die openbare, kommersiële of nie-winsgewende sektore.

Data beskikbaarheidsverklaring

Die deel van data is nie van toepassing op hierdie artikel nie, aangesien geen nuwe data in hierdie studie geskep of ontleed is nie.

Vrywaring

Die sienings en menings wat in hierdie artikel uitgedruk word, is dié van die outeurs en weerspieël nie noodwendig die amptelike beleid of posisie van enige geaffilieerde agentskap van die outeurs nie.

Literatuurverwysings

Bachman, L.F. & Palmer, A.S., 1996, *Language testing in practice: Designing and developing useful language tests*, Oxford University Press, Oxford.

Bizzel, P., 1992, *Academic discourse and critical consciousness*, University of Pittsburgh Press, Pittsburgh.

Blanton, L.L., 1994, 'Discourse, artefacts and the Ozarks: Understanding academic literacy', *Journal of Second Language Writing* 3(1), 1–16. Herdruk (as Hoofstuk 17: 219–235) in Zamel, V. & Spack, R. (eds.), 1998.

BusinessTech (staff writer), 13 November 2019, *South Africa's university pass rate shocker*, viewed 17 Oktober 2020, from <https://businesstech.co.za/news/government/353575/south-africas-university-pass-rate-shocker/>

Butler, H.G., 2007, 'A framework for course design in academic writing for tertiary education', Unpublished PhD thesis, University of Pretoria, Pretoria.

Carstens, A. & Eiselen, R., 201, 'Designing a South African Multilingual Learner Corpus of Academic Texts (SAMuLCAT)', *Language Matters* 50(1), 64–83.

Charles, M., Pecorari, D. & Hunston, S., 2009, *Academic writing at the interface of corpus and discourse*, Continuum, New York, NY.

Coxhead, A., 2000, 'A new academic word list', *TESOL Quarterly* 34(2), 213–238. <https://doi.org/10.2307/3587951>

De Wachter, L., D'Hertefelt, M. & Heeren, J., 2016, *De digitale Schrijfhulp Nederlands: Een processgeoriënteerde schrijfhulp ter bevordering van schrijfvaardigheid in het hoger onderwijs*, viewed 20 Oktober 2019, from <http://hdl.handle.net/1887/38707>

Department of Science and Technology, 2016, *South African research infrastructure roadmap*, Department of Science and Technology, Pretoria.

Doughty, C. & Pica, T., 1986, "Information Gap" tasks: Do they facilitate second language acquisition?", *TESOL Quarterly* 20(2), 305–325. <https://doi.org/10.2307/3586546>

Duff, P.A. & Hornberger, N., 2010, *Language socialization: Encyclopedia of language and education*, vol. 8, Springer, New York.

Flowerdew, L., 2001, 'The exploitation of small learner corpora in EAP materials design', in M. Ghadessy & R. Roseberry (eds.), *Small corpus studies and ELT*, pp. 363–379, Benjamins, Amsterdam.

Flowerdew, L., 2009, 'Applying corpus linguistics to pedagogy: A critical evaluation', *International Journal of Corpus Linguistics* 14(3), 393–417. <https://doi.org/10.1075/ijcl.14.3.05flo>

Flowerdew, J. & Peacock, M. (eds.), 2001a, *Research perspectives on English for academic purposes*, pp. 8–24, Cambridge University Press, Cambridge.

Flowerdew, J. & Peacock, M., 2001b, 'Issues in EAP: A preliminary perspective', in J. Flowerdew & M. Peacock (eds.), *Research perspectives on English for academic purposes*, pp. 8–24, Cambridge University Press, Cambridge.

Gee, J.P., 1990, *Social linguistics and literacies: Ideology in discourses*, Falmer Press, London.

Gee, J.P., 1998, 'What is literacy?' (pp. 51–59) in V. Zamel & R. Spack (eds.), 1998: (reprint of the 1987 article), *Teaching and Learning: The Journal of Natural Enquiry* 2, 3–11.

Gee, J.P., 2000, 'The new literacy studies: From "socially situated" to the work of the social', in D. Barton, M. Hamilton & R. Ivanic (eds.), *Situated literacies: Reading and writing in context*, pp. 180–196, Routledge, London.

Ghadessy, M. & Roseberry, R. (eds.), *Small corpus studies and ELT*, pp. 363–379, Benjamins, Amsterdam.

Gilquin, G., Granger, S. & Paquot, M., 2007, 'Learner corpora: The missing link in EAP pedagogy', *Journal of English for Academic Purposes* 64, 319–335. <https://doi.org/10.1016/j.jeap.2007.09.007>

Huang, L.-S., 2017, 'Taking stock of corpus-based instruction in teaching English as an international language', *RELC Journal* 49(3), 381–401. <https://doi.org/10.1177/0033688217698294>

Hyland, J., 2009, 'Corpus informed discourse analysis: The case of academic engagement', in M. Charles, D. Pecorari & S. Hunston (eds.), *Academic writing at the interface of corpus and discourse*, pp. 110–128, Continuum, New York, NY.

Hyland, K., 2004, *Genre and second language writing*, University of Michigan Press, Ann Arbor, MI.

Hyland, K., 2008, 'Genre and academic writing in the disciplines', *Language Teaching* 41(4), 543–562. <https://doi.org/10.1017/S0261444808005235>

Hyland, K. & Hyland, F., 2006, 'Feedback on second language students' writing', *Language Teaching* 39(2), 83–101. <https://doi.org/10.1017/S0261444806003399>

Instituut voor levende talen, 2020, viewed 15 November 2019, from <https://ilt.kuleuven.be/english/>

Inter-Institutional Centre for Language Development and Assessment (ICELDA), 2020, viewed 15 Mei 2020, from <https://icelda.com/>

Krishnamurthy, R. & Kosem, I., 2007, 'Issues in creating a corpus for EAP pedagogy and research', *Journal of English for Academic Purposes* 6(4), 356–373. <https://doi.org/10.1016/j.jeap.2007.09.003>

Lasagabaster, D., 2008, 'Foreign language competence in content and language integrated courses', *The Open Applied Linguistics Journal* 1, 30–41. <https://doi.org/10.2174/1874913500801010030>

Leech, G., 1992, 'Corpora and theories of linguistic performance', in J. Svartvik (ed.), *Directions in corpus linguistics*, pp. 102–123, Mouton/De Gruyter, Berlin.

Louw, H., 2006, 'Standardising written feedback on L2 student writing', Master's dissertation, North-West University (Potchefstroom Campus).

Louw, H., 2007, 'Moving to more than editing: Standardised feedback in practice', *Ensovoort* 11(2), 83–104.

Louw, H., 2011, 'MarkWrite: Standardised feedback on ESL student writing via a computerised marking interface', Thesis – PhD, NWU, Potchefstroom.

- McEnery, T. & Xiao, R., 2011, 'What corpora can offer in language teaching and learning', in E. Hinkel (ed.), *Handbook of research in second language teaching and learning: Volume II*, pp. 364–380, Routledge, New York, NY.
- McEnery, T., Xiao, R. & Tono, Y., 2006, *Corpus survey*, viewed 16 March 2020, from <http://bowland-files.lancs.ac.uk/corplang/cbts/corpora.asp>
- Meihuizen, E.M.M., 2019, 'Formative assessment in academic writing: Integrating online feedback within the broader teaching-learning community', *Journal for Language Teaching* 53(1), 195–215. <https://doi.org/10.4314/jlt.v53i1.9>
- Oller, J.W. & Richards, J.C. (eds.), 1973, *Focus on the learner: Pragmatic perspectives for the language teacher*, Newbury House, Rowley, MA.
- Roos, H.J., 1991, 'A syntactic error analysis of written work of students at Vista University: Implications for remediation', Dissertation – MA, Randse Afrikaanse Universiteit.
- Skehan, P., 1988, 'Language testing, Part 1', *Language teaching*, 21(4), 11–221. <https://doi.org/10.1017/S0261444800005218>
- South African Centre for Digital Language Resources (SADiLaR), 2020, viewed 15 Mei 2020, from <https://www.sadilar.org/index.php/en/>
- Spolsky, B., 1973, 'What does it mean to know a language, or how do you get someone to perform his competence?', in J.W. Oller & J.C. Richards (eds.), pp. 164–176.
- Stander, M., 2000, 'Syntactic structures of Afrikaans second language speakers', *Journal for Language Teaching* 34(1), 6.
- Svartvik, J. (ed.), 1992, *Directions in Corpus Linguistics*, pp. 102–123, Mouton/De Gruyter, Berlin.
- Swales, J. & Feak, C., 2000, *Academic writing for graduate students: Essential tasks and skills*, University of Michigan Press, Ann Arbor, MI.
- Tribus, A.C., 2017, 'The communicative functions of language: An exploration of Roman Jakobson's theory in TESOL', *MA TESOL Collection* 723, viewed 20 October 2020. from https://digitalcollections.sit.edu/ipp_collection/723/
- Van de Poel, K. & Gasiorek, J., 2012, 'Effects of an efficacy-focused approach to academic writing on students' perceptions of themselves as writers', *Journal of English for Academic Purposes* 11(4), 194–303. <https://doi.org/10.1016/j.jeap.2012.07.003>
- Van Dyk, T.J., 2010, 'Konstitutiewe voorwaardes vir die ontwerp van 'n toets van akademiese geletterdheid', Ongepubliseerde PhD-tesis, Universiteit van die Vrystaat, Bloemfontein.
- Van Dyk, T.J., 2015, 'Tried and tested: Academic literacy tests as predictors of academic success', *Tijdschrift voor Taalbeheersing* 37(2), 159–186.
- Van Dyk, T.J. & Van de Poel, K., 2013, 'Towards a responsible agenda for academic literacy development: Considerations that will benefit students and society', *Journal for Language Teaching* 47(2), 43–70. <https://doi.org/10.4314/jlt.v47i2.3>
- Van Dyk, T.J. & Weideman, A.J., 2004, 'Switching constructs: On the selection of an appropriate blueprint for academic literacy assessment', *Journal for Language Teaching* 38(1), 1–13. <https://doi.org/10.4314/jlt.v38i1.6024>
- Weideman, A.J., 2003, 'Assessing and developing academic literacy', *Per Linguam* 19(1&2), 55–65. <https://doi.org/10.5785/19-1-89>
- Xue, G. & Nation, I.S.P., 1984, 'A university word list', *Language Learning and Communication* 3(2), 215–229.
- Yoon, H., 2008, 'More than a linguistic reference: The influence of corpus technology on L2 academic writing', *Language Learning & Technology* 12(2), 31–48.
- Zamel, V. & Spack, R. (eds.), 1998, *Negotiating academic literacies: Teaching and learning across languages and cultures*, Lawrence Erlbaum, Mahwah, NJ.