CURRICULUM VITAE
Gideon (Jozua) Kotzé, PhD

Woning: Pretoria, South Africa
E-pos: gidi8ster@gmail.com
Webwerf: www.gideonkotze.co.za

PERSOONLIKE INLIGTING

Nasionaliteit: Suid-Afrikaans
Geboortedatum: 9 Junie 1981
Tale magtig: In dalende volgorde van vaardigheid: Afrikaans (moedertaal), Engels, Nederlands, Duits en Frans. Beperkte kennis van isiXhosa (hoërskoolopleiding) en Japannees (gesproketaalkursus).

OPSOMMING

Ek het ’n PhD in rekenaarlinguistiek van die Rijksuniversiteit Groningen (2013), met 18 navorsingspublikasies. Ek het ’n sterk belangstelling in die ontwikkeling van taalhulpbronne en -tegnologieë, veral met die fokus op Suid-Afrikaanse tale. My navorsing by die Universiteit van Suid-Afrika (Unisa; 2014-2019) het binne die volgende projekte plaasgevind: die ontwikkeling en semi-outomatiese skepping van geannoteerde XML-tekskorpora vir Suid-Afrikaanse tale, die outomatiese digitalisering van fisiese dokumente met ’n klem op hoë kwaliteit, sowel as statistiese en neurale masjienvertaling. My proefskrif handel oor die outomatiese belyning van sogenaamde parallelle (sintaktiese) boombanke wat ten doel het om opleidingsdata te verskaf vir sintaksisgebaseerde masjienvertaling. Leksikografiese ervaring sluit my betrokkenheid as hoof van die Afrikaanse woordnetprojek by die Sentrum vir Tekstegnologie (2006-7), sowel as projekwerk tydens my Meestersgraad in Linguistiek (Leksikologie en Terminologie) by die Vrije Universiteit in Amsterdam (2005-6) in. By Unisa was ek ook betrokke by die aanbieding van gereelde werkswinkels wat deelnemers help om Wikipedia-inhoud te ontwikkel vir Afrikatale. Ten slotte het ek ook ervaring van studentebegeleiding, dosering en die skryf van subsidievoorstelle.

WERKSONDERVINDING

2020 Projekwerk vir die “South African Centre for Digital Language Resources” (SADiLaR) deur die Departement van Afrikatale by die Universit van Suid-Afrika, Pretoria, as onafhanklike kontrakteur. Dit behels die ontwikkeling van ’n sagteware-omgewing wat navrae stuur na ’n vanlyn-weergawe van die SQL-databasis van die Afrika-“wordnet” wat inligting inwin, tabelle opdateer en statistieke genereer, en is gebaseer op data van sigblaaie. Die doel hiervan is die toepassing van kwaliteitsbeheer en om die uitbreiding van die wordnet te versnel.

2019-2020 Onafhanklike kontrakteur vir die Suid-Afrikaanse Akademie vir Wetenskap en Kuns (SAAWK) vir die skryf van ’n handleiding in Afrikaans om lesers by te staan met die redigering en gebruik van die Afrikaanse Wikipedia. Prof. Laurette Pretorius is mede-outeur.

2016-2019 Rekenaarlinguis (senior navorser) by die Akademie vir Afrikatale en Wetenskap (Academy for African Languages and Science; AALS), Universiteit van Suid-Afrika (Unisa), onder Prof. Laurette Pretorius, en in die laaste paar maande, Me. Nozibele Nomdebevana. Projekaktiwiteite sluit die volgende in: die implementasie van ’n

raamwerk vir TEI P5-korpusskepping en kwaliteitsbeheer van brondokumente (Afrikaans en Engels), die skepping van ’n pyplyn vir die digitalisering van institusionele dokumente van Unisa, masjienvertaaleksperimente (statisties en neuraal; Zoeloe/Engels en Tswana/Engels), sowel as die versameling, samevoeging, skoonmaak en filtrering van korpora en parallelle korpora. Terwyl ek by Unisa gewerk het was ek ook verantwoordelik vir die volgende: die onderhoud van die AALS-webwerf,[1] die mede-ontwikkeling van ’n webwerf vir die internasionale “Grammatical Framework”-somerskool in Stellenbosch (2018),[2] sowel as die bydrae tot en die aanbieding van praatjies by maandelikse werkswinkels waar AALS deelnemers gehelp het om in Suid-Afrikaanse tale inhoud vir Wikipedia te ontwikkel. Vanaf 2014 het ek ook by verskeie geleenthede praatjies aangebied (sien Praatjies en Aanbiedings).

2014-2016 Postdoktorale genoot by Unisa onder prof. Laurette Pretorius. In hierdie periode het my navorsing gefokus op die statistiese masjienvertaling van Zoeloe na Engels, sowel as die ontwerp en implementasie van ’n pyplyn vir die bogenoemde meertalige korpus van institusionele dokumente van Unisa. In 2015 het ek twee Honneursstudente in ’n jaarmodule begelei (“Honours Research Report” - HRCOS82). Die studente het hulle grade in 2016 voltooi (BSc Honneurs in Rekenaarkunde en BCom Honneurs in Besigheidsinformatika).

2012 In hierdie periode, wat die laaste jaar van my PhD was, het ek gewerk aan die ontwerp en implementering van ’n sogenaamde transformasie-gebaseerde sisteem wat belynings tussen sintaktiese frasestruktuurbome outomaties skep of korrigeer. Die doel van hierdie projek was die skepping van sogenaamde parallelle boombanke op groot skaal wat dan gebruik word in sintaksisgebaseerde masjienvertaling, soos met die “PaCo-MT”-projek (sien onder). ’n Artikel oor hierdie werk is gepubliseer in die tydskrif “Computational Linguistics in the Netherlands” (CLIN) in 2012. Hierdie werk vorm deel van die navorsing vir my PhD-graad (sien Opleiding).

2009 By die Rijksuniversiteit Groningen het ek ’n kursus genaamd “Tekstmanipulatie” aangebied op voorgraadse vlak saam met Dr. Barbara Plank.

2008-2011 In hierdie periode het ek aan die masjienvertaalprojek “Parse and Corpus-Based Machine Translation” (PaCo-MT) gewerk saam met die Katolieke Universiteit Leuven, België, en die maatskappy OneLiner bvba in Sint Niklaas, België. Die ander lede van die konsortium het bestaan uit Frank van Eynde, Vincent Vandeghinste, Jörg Tiedemann, Joachim van den Bogaert en Koen Desmet. My belangrikste taak was om oplossings te vind of te ontwikkel vir die skepping van groot parallelle boombanke wat gebruik kan word in die masjienvertaalsisteem. Ek het hoofsaaklik tekenvereenduidiging (“tokenisation”), sinsbelyning, woordbelyning, sintaktiese ontleding en boom-tot-boom-belyning (“tree alignment”) toegepas. Die betrokke taalpare was Nederlands/Engels en Nederlands/Frans, in beide rigtings.

2008 In hierdie periode het ek kontrakwerk verrig vir die Sentrum vir Tekstegnologie (CTexT) in Potchefstroom. My hooftaak was die uitbreiding van die Afrikaanse “woordnet” (sien onder) na ’n totaal van 10 068 inskrywings, waar Engelse inskrywings hoofsaaklik in Afrikaans vertaal is. Die taak het ook datamanipulasie in die programmeertaal Perl vereis.

2006-2007 Tydelike pos as data-ontwikkelaar en navorser by CTexT, met Prof. Gerhard van Huyssteen as hoof van die sentrum. Aktwiteite sluit die volgende in: morfologiese analise en kwaliteitsbeheer, datamanipulasie deur middel van programmering (in Perl), die ontwikkeling van databasisse (spesifiek ALEXANDER - “Afrikaans Lexicon and Annotated Database for Engineering and Research”, sowel as die Afrikaanse

1 www.unisa.ac.za/aals
2 school.grammaticalframework.org/2018

	woordnet), taalkundige versorging en kwaliteitsbeheer van leksika, korpusanalise en vertaling.
2006	Internskap by die Vrije Universiteit in Amsterdam as deel van my Meestersgraad in Leksikologie en Terminologie. Ek het gewerk aan ’n projek genaamd Cornetto (“Combinatorial and Relational Network as Toolkit for Dutch Language Technology”), ’n elektroniese leksikale databasis in Nederlands vir gebruik in mensetaaltegnologie. Ek het onder andere semantiese annotasie en datamanipulasie deur middel van programmering verrig.
2002-2004	Rekenaarassistent by die voormalige Universiteit van Port Elizabeth (tans die Nelson Mandela-Universiteit) in rekenaarlaboratoriums vir studente.

REKENAARVAARDIGHEDE

Ek beskik oor ervaring van Linux-sisteme (Ubuntu, Mint, Fedora), MacOS (Lion tot en met El Capitan), sowel as Windows. Die laaste paar jaar het ek meestal op Linux geprogrammeer.

Programmering
Python (hoofkeuse vir projekwerk vanaf omtrent 2017), Perl (hoofkeuse vir die PaCo-MT-projek en my PhD-navorsing), Bash. Meer beperkte kennis van C++ en Java (kursusse op voorgraadse vlak), R en Turbo Pascal. Ek het kennis verwerf oor XHTML, CSS, JavaScript, VBScript en ASP as deel van ’n webtegnologiekursus op voorgraadse vlak. Meer onlangs het ek kennis verwerf oor HTML 5 en CSS 3 en ook hierdie tegnologieë by die werk toegepas. Ek het ervaring van databasisse (Microsoft Access, SQL). As ’n manier om data te verteenwoordig het ek baie ervaring van die gebruik van XML, XML-ontleders (Python se lxml.etree en Perl se XML::Twig), sowel as ’n bietjie gebruik van XPath en XSLT. Ten slotte het ek ook ervaring van die weergawebeheersagteware Git en gebruik dit op ’n daaglikse basis.

Rekenaarlinguistieksagteware
NLTK-raamwerk (sinsverdeling, woordsoortetikettering, lemmatisering, taalherkenning), die statistiesemasjienvertaalsisteem Moses, die neuralemasjienvertaalsisteem OpenNMT, meer beperkte gebruik van sintaksgebaseerde masjienvertaling (PaCo-MT), belyningsagteware (GIZA++, Fast Align), handmatige belyningsagteware (InterText, Stockholm TreeAligner), woordsoortetiketteerders (bv. TreeTagger, Stanford, NCHLT[3]), sintaktiese ontleders (Stanford, Berkeley, Alpino, MaltParser), die sintaktiese boombelyners Lingua-Align, Dublin Sub-Tree Aligner and TBLign (eie sagteware), ander sagteware soos sinsverdelers (Moses se verdeler, Punkt) en sinsbelyners (Gale & Church, Hunalign), en ten slotte die dokumentontrekking, -ontleding en -opmaakprogrammatuur GROBID (2018-werkswinkel).

Taalkundige sagteware
BlackLab, WhiteLab, MultiTerm, WordSmith, Alchemist, Stockholm TreeAligner, ParaConc.

Teksproduksie en relevante vaardighede
Microsoft Office (Excel, Word, Access, Powerpoint), OpenOffice en LibreOffice (teks-, sigblad- en voordragsagteware), sowel as LaTeX vir akademiese teksproduksie. Ek het ook ervaring van verskeie redigeerders soos Atom, Sublime Text, Emacs, Nano, XML Copy Editor, Kate, KWrite, TeXShop, TexMaker, vir beide teksproduksie en programmering.

Ander sagteware
Beeldverwerking en optiesekarakterherkenning (OCR)-sisteme: ImageMagick, Unpaper, Tesseract, Calamari, OCRopus, OCRFeeder, gscan2pdf. Ek het almal behalwe die laaste twee in ’n digitaliseringspyplyn, met die doel om institusionele dokumente van Unisa te verwerk, geïntegreer (2017-2019).

3 Dit het ook die vermoë om tekenverdeling (“tokenisation”) en frasegroepering (“phrase chunking”) toe te pas en sinsgrense te vind, en annoteer ook lemmas en naamentiteite.

OPLEIDING

Instellings

- Rijksuniversiteit Groningen, Nederland (2008-2012) (PhD)
- Vrije Universiteit in Amsterdam, Nederland (2005-2007) (Meestersgraad)
- Die voormalige Universiteit van Port Elizabeth, tans die Nelson Mandela-Universiteit (B.A. en B.A. Honneurs)
- Eberhard Karls-Universiteit Tübingen, Duitsland (2004) (uitruilprogram vir studente)
- Universiteit van Suid-Afrika (Algemene Linguistiek I) (2001)

Grade en kursusse

2011 Europese Somerskool in Logika, Taal en Inligting (European Summer School in Logic, Language and Information; ESSLLI) in Ljubljana, Slowenië, bygewoon.

2011 'n Masjienleerkursus by die Rijksuniversiteit Groningen bygewoon, wat gelei het tot 'n artikel genaamd "Authorship attribution of weblog texts: A comparative study".

2010 ESSLLI-somerskool in Kopenhagen, Denemarke, bygewoon.

2010 Die 4de "Masjienvertaalmarathon" by die Stadsuniversiteit Dublin bygewoon.

2009 ESSLLI-somerskool in Bordeaux, Frankryk, bygewoon.

2008-2013 **Doktor in die Filosofie (Rekenaarlinguistiek)**

2013 Doktorale proefskrif voltooi en aanvaar vir openbare verdediging: *Complementary approaches to tree alignment: Combining statistical and rule-based methods.* In hierdie werk word die belyning van sintaktiese frasestrukture van vertaalde tekste ondersoek met die doel om sogenaamde *parallelle boombanke* te skep, waarvan die hooftoepassing sintaksisgebaseerde masjienvertaling is. Alhoewel die effek op die verrigting van dergelike vertaling nog tekort skiet, word daar gewys dat verskeie nuwe benaderings tot die belyningsprobleem suksesvol gekombineer kan word om 'n hoëkwaliteit-bron te skep.

2005-2007 **Magister Artium in Taalkunde (Leksikologie en Terminologie)**

2007 Skripsie voltooi: *Building a WordNet for Afrikaans: Preliminary research in the form of an inquiry into the feasibility and optimal methodology for the development of a wordnet database*. Dit is 'n voorlopige studie oor 'n voorgestelde projek by die Noordwes-Universiteit in Potchefstroom, Suid-Afrika. Dit handel oor die ontwerp en bou van 'n Afrikaanse *woordnet* – 'n semantiese netwerk wat as 'n elektroniese leksikale databasis gestoor word en vir leksikografiese en mensetaaltegnologiese gebruik. Die projek is befonds deur die destydse Suid-Afrikaanse Departement van Wetenskap en Tegnologie.

2005-2006 Kursusse by die Vrije Universiteit in Amsterdam, Nederland (name is in Nederlands): "Lexicografische en Terminologische Tools", "Bilinguale Lexicografie", "Werkcollege Lexicologie", "Werkcollege Terminologie", "Lexicale Semantiek", "Computationele Lexicologie" (geen eksamen), "Relationele Modellen", "Vaktaal".

2004-2005 **B.A. Honneurs in Toegepaste Taalstudies (Vertaalstudies)** (*cum laude*)

2004 Kursusse by die destydse Universiteit van Port Elizabeth:
Inleiding tot Algemene Linguistiek, Toegepaste Linguistiek, Vertaalteorie.
Kursusse by die Eberhard Karls-Universiteit Tübingen, Duitsland: Rekenaarlinguistiek (Sintaktiese Analise en Programmering in Java (Datastrukture)), Gevorderde Duits, Gesproke Duits.

2000-2003 **B.A. in Toegepaste Taalstudies** (*cum laude*)

2003 Kursusse: Afrikaans III (hoofsaaklik Linguistiek en Literatuur), Duits III, Frans III en Musiektegnologie.

2001-2002 Kursusse: Afrikaans I en II, Duits I en II, en Frans I en II. Bykomende kursusse: Programmering in C++, Programmering in Java (Algoritmiek en Datastrukture,

Webtegnologie, Wiskunde (Algebra en Calculus), Inligtingstelsels en Rekenaarargitektuur.

2000 Kursusse: Afrikaans, Nederlands, Inleiding tot die Universiteitspraktyk en Rekenaargeletterdheid.

AKADEMIESE PRESTASIES EN ERVARING

Publikasies

2020

Kapanadze, O., Kotzé, G. en Hanneforth, T. 2020. Building Resources for Georgian Treebanking-based NLP (ingedien). In: *13th International Tbilisi Symposium on Language, Logic and Computation, TbiLLC 2019*, Batumi, Georgië, 16-20 September, 2019. Revised Selected Papers.

Kotzé, G. en Wolff, F. Aanvaar. Exchanging image processing and OCR components in a Setswana digitisation pipeline. *South African Computer Journal*. South African Institute of Computer Scientists and Information Technologists (SAICSIT).

2019

Hanneforth, T., Kapanadze, O. en Kotzé, G. Applying Computer Technologies to the Georgian Language: From a Treebank to a Syntactic Parser [Long abstract]. Aangebied by *The Thirteenth International Tbilisi Symposium on Language, Logic and Computation*, Batumi, Georgië, 16-20 September 2019. URL: http://events.illc.uva.nl/Tbilisi/Tbilisi2019/uploaded_files/inlineitem/Kapanadze.pdf.

2017

Kotzé, G., Vandeghinste, V., Martens, S. & Tiedemann, J. 2017. Large aligned treebanks for syntax-based machine translation. *Language Resources and Evaluation*. Vol. 51(2). Springer. doi:10.1007/s10579-016-9369-0. URL: http://link.springer.com/article/10.1007/s10579-016-9369-0.

Kotzé, G. & Wolff, F. 2017. Developing and evaluating a pipeline for Setswana OCR. *Proceedings of the 2017 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech)*, bl. 236-241. IEEE. URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8261154.

2016

Kotzé, G. 2016. Refining semi-automatic parallel corpus creation for Zulu to English statistical machine translation. In: *Proceedings of the 2016 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech)*, bl. 48-53. IEEE. http://ieeexplore.ieee.org/document/7813168/.

2015

Kotzé, G. & Wolff, F. 2015. Syllabification and parameter optimisation in Zulu to English machine translation. *South African Computer Journal*. No. 57, bl. 1-23. South African Institute of Computer Scientists and Information Technologists (SAICSIT). http://dx.doi.org/10.18489/sacj.v0i57.323

2014

Wolff, F. & Kotzé, G. 2014. Experiments with syllable-based Zulu-English machine translation. In Puttkammer, M. en Eiselen, R. (reds.): *Proceedings of the 2014 PRASA, RobMech and AfLaT International Joint Symposium*. bl. 217-222. Kaapstad, Suid-Afrika.

Kotzé, G. & Wolff, F. 2014. Experiments with syllable-based English-Zulu alignment. In *Proceedings of the SaLTMiL Workshop on free/open-source language resources for the machine translation of less-resourced languages* (by LREC 2014), bl. 7-11, Reykjavík, Ysland.

2013

Kotzé, G. 2013. *Complementary Approaches to Tree Alignment: Combining Statistical and Rule-Based Methods*. PhD-proefskrif. Rijksuniversiteit Groningen.

Vandeghinste, V., Martens, S., Kotzé, G., Tiedemann, J., Van den Bogaert, J., De Smet, K., Van Eynde, F., & Van Noord, G. 2013. Parse and Corpus-based Machine Translation. In *Peter Spyns and Jan Odijk (reds.): Essential Speech and Language Technology for Dutch*. bl. 305-319. Springer.
https://link.springer.com/chapter/10.1007%2F978-3-642-30910-6_17.

2012

Kotzé, G. 2012. Transformation-based tree-to-tree alignment. In *Computational Linguistics in the Netherlands Journal*. Vol. 2, bl. 71-96.

Kotzé, G., Vandeghinste, V., Martens, S. & Tiedemann, J. 2012. Large aligned treebanks for syntax-based machine translation. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, bl. 467-473, Istanboel, Turkye.

2011

Kotzé, G. 2011. Finding Statistically Motivated Features Influencing Subtree Alignment Performance. In *Bolette Sandford Pedersen, Gunta Nešpore and Inguna Skadiņa (eds.): Proceedings of the 18th Nordic Conference of Computational Linguistics NODALIDA 2011*, May 11-13, Riga, Letland. NEALT Proceedings Series, Vol. 11, bl. 332-335. Tartu: Tartu University Library.

Kotzé, G. 2011. Improving syntactic tree alignment through rule-based error correction. In *Proceedings of ESSLLI 2011 Student Session*, bl. 122-127, Ljoebljana, Slowenië.

Kotzé, G. 2011. Rule-induced error correction of aligned parallel treebanks. In *Proceedings of the International Conference "Corpus Linguistics - 2011"*, bl. 35-40, Sint Pietersburg, Rusland.

Vandeghinste, V., Van den Bogaert, J., Martens, S., & Kotzé, G. 2011. PaCo-MT: Parse and Corpus-based Machine Translation. In *Forcada, M.L., Depraetere, H., and Vandeghinste, V. (reds.): Proceedings of the 15th International Conference of the European Association for Machine Translation*, bl. 347. Leuven, België.

2009

Tiedemann, J. & Kotzé, G. 2009. A Discriminative Approach to Tree Alignment. In *Ilisei, I., Pekar, V. en Bernardini, S. (reds.): Proceedings of the International Workshop on Natural Language Processing Methods and Corpora in Translation, Lexicography and Language Learning* (in verband met RANLP'09), bl. 33-39. Borovets, Bulgarye.

Tiedemann, J. & Kotzé, G. 2009. Building a Large Machine-Aligned Parallel Treebank. In *Passarotti, M., Przepiórkowski, A., Raynaud, S. en Van Eynde, F. (reds.): Proceedings of the 8th*

International Workshop on Treebanks and Linguistic Theories (TLT'08), bl. 197-208. Milaan, Italië.

2008

Kotzé, G. 2008. Development of an Afrikaans wordnet: methodology and integration / Ontwikkeling van ’n Afrikaanse woordnet: metodologie en integrasie. *Literator: Journal of Literary Criticism, Comparative Linguistics and Literary Studies: Human language technology for South African languages: Special Issue 1*, Volume 29(1). bl. 163-184.

2006

Kotzé, G. 2006. *Building a WordNet for Afrikaans: Preliminary research in the form of an inquiry into the feasibility and optimal methodology for the development of a wordnet database*. Skripsie vir Meestersgraad. Vrije Universiteit Amsterdam.

Beoordeelde artikels

- 1 artikel vir die verrigtinge van die “2017 Pattern Recognition Association of South Africa and Robotics and Mechatronics”-konferensie (PRASA-RobMech)
- 1 artikel vir die verrigtinge van die “2019 SAUPEC / ROBMECH / PRASA”-konferensie
- 1 artikel vir insluiting by ’n spesiale uitgawe van die Suid-Afrikaanse Rekenaartydskrif, bestaande uit ’n versameling van uitgebreide artikels wat by die Forum vir Navorsing in Kunsmatige Intelligensie in Desember 2019 in Kaapstad aangebied is.
- (nog nie ingedien nie) 1 artikel vir insluiting by die “Language Resources and Evaluation”-tydskrif (2020), ’n Springer-publikasie.

Studentebegeleiding

In 2015 het ek by Unisa twee Honneurstudente in ’n jaarmodule begelei (“Honours Research Report” - HRCOS82). Die studente het hulle grade in 2016 voltooi (BSc Honneurs in Rekenaarkunde en BCom Honneurs in Besigheidsinformatika).

Kursusse aangebied

In 2009 het ek ’n kursus op voorgraadse vlak by die Rijksuniversiteit Groningen aangebied, getiteld “Tekstmanipulatie”, saam met Dr. Barbara Plank.

Leierskapsposisies

2014-2019 Akademie vir Afrikatale en Wetenskap (AALS), Universiteit van Suid-Afrika (Unisa), by die hoofkampus op Muckleneukrand: leier van digitaliserings- en masjienvertaalprojekte.

2006-2007 Sentrum vir Tekstegnologie (CTexT) by die Potchefstroomkampus van die Noordwes-Universiteit:

- Projekleier van die leksikale databasis ALEXANDER (“Afrikaans Lexicon and Annotated Database for Engineering and Research”) (2006)

- Projekleier van die Afrikaanse woordnet (2007 – sien “OPLEIDING”)

PRAATJIES EN AANBIEDINGS

- Twee praatjies by die SAAWK aangebied oor die gebruik van die Afrikaanse Wikipedia vir skryf- en redigeerwerk. Dit het in Pretoria op 30 Januarie 2020 plaasgevind.
- Praatjie en demonstrasie by ‘n tweedaagse Unisa-werksessie wat die voorlopige bevindinge van geoutomatiseerde kwaliteitsbeheer vir die Afrika-wordnet voorgelê het. Dit het op 28 en 29 Januarie 2020 in die Burgers Park-hotel, Pretoria, plaasgevind.
- Wikipedia-werkswinkel georganiseer en aangebied by die Potchefstroomkampus van die Noordwes-Universiteit, die gasheer van die geleentheid, op 3 April 2019, verteenwoordigend van AALS, Unisa, saam met Dr. Friedel Wolff en Me. Motswalle Kanyane. Die Suid-Afrikaanse Sentrum vir Digitale Taalhulpbronne (SADiLaR) was die borg.
- Praatjie op 6 September 2018 by die “DEASA 2018”-konferensie by Unisa in Pretoria, getitled “A freely accessible, multilingual corpus for use in South African open learning environments”.
- Praatjie op 23 September 2016 by ’n Unisa-seminaar in Pretoria op die geleentheid van Internasionale Vertaaldag, getiteld “The power of the text corpus: A Unisa perspective”.
- Gesamentlike aanbieding met Friedel Wolff op 20 Augustus 2015 by die Nasionale Konferensie oor Meertaligheid in Hoër Onderwys (National Conference on Multilingualism in Higher Education) by Unisa in Pretoria, getiteld “Machine Translation at Unisa”. Ek was ook die mede-outeur van ’n aanbieding by dieselfde geleentheid met die titel “Building language resources from institutional content in Higher Education”.
- Praatjie op 26 Junie 2015 by die gesamentlike konferensie van LSSA/SAALA/SAALT (“Linguistic Society of Southern Africa” (LSSA), die “South African Applied Linguistics Association” (SAALA) en die “South African Association for Language Teaching” (SAALT)) by die Noordwes-Universiteit in Potchefstroom, getiteld “Applying English morphological segmentation to syllable-based Zulu-to-English statistical machine translation”.
- Praatjie op 22 April 2015 by Unisa se plaaslike “Machine Translation”-seminaar, georganiseer deur AALS en die College vir Nagraadse Studies (College of Graduate Studies of CGS). Die titel was “Rule-based syntactic tree alignment of parallel translated sentences” en is gebaseer op ’n versameling hoofstukke van my PhD-proefskrif.
- Praatjie op 9 Julie 2014 by Unisa se plaaslike “Sustainable Multilingualism Seminar”, georganiseer deur AALS en CGS. Die titel was “Experiments with syllable-based English-Zulu alignment” en is gebaseer op een van my artikels met dieselfde naam.
- Praatjie op 27 Mei 2014 by die “SaLTMiL Workshop on free/open-source language resources for the machine translation of less-resourced languages” by die “Language Resources and Evaluation”-konferensie (LREC 2014) in Reykjavík, Ysland.
- Verdediging van doktorale proefskrif by die Rijksuniversiteit Groningen op 24 Junie 2013, wat gelei het tot die toekenning van ’n PhD-graad.
- Plakkaataanbieding by die studentesessie van die Europese Somerskool in Logika, Taal en Inligting (European Summer School in Logic, Language and Information; ESSLLI 2011) in Ljoebljana, Slowenië. Die titel is “Improving syntactic tree alignment through rule-based error correction”. (gepubliseer)
- Praatjie by “Corpus Linguistics 2011” in Sint Pietersburg, Rusland. The title is “Rule-induced error correction of aligned parallel treebanks”. (gepubliseer)
- Plakkaataanbieding by die Nordiese Konferensie oor Rekenaarlinguistiek (Nordic Conference on Computational Linguistics; NoDaLiDa'19) in Riga, Letland, op 11-13 Mei 2011. Titel: “Finding Statistically Motivated Features Influencing Subtree Alignment Performance”. (gepubliseer)

- Praatjie by die “Computational Linguistics in the Netherlands”-konferensie (CLIN 2011) conference in Gent, België. Die titel is dieselfde as bogenoemde: “Finding Statistically Motivated Features Influencing Subtree Alignment Performance”.
- Uitgenooi in 2010 deur Prof. Dr. Martin Volk as ’n spreker by ’n plaaslike kollokwium by die Instituut vir Rekenaarlinguistiek (Institute for Computational Linguistics) by die Universiteit van Zürich, Switserland. Die onderwerp was die boom-tot-boom-belyningsagteware Lingua-Align wat deur Jörg Tiedemann ontwikkel is. Ek het die konferensieartikel wat deur Tiedemann geskryf en by LREC 2010 in Malta voorgedra is, hier aangebied.
- Praatjie by “TABU Dag 2010” (TABU: “taalbulletin”) in Groningen, getiteld “Supervised tree alignment for syntax-based machine translation”.
- Plakkaataanbieding by “Computational Linguistics in the Netherlands” (CLIN 2010) in Utrecht, Nederland. Title: “Discriminative parallel treebank alignment for syntax-based machine translation”.
- Konferensiepraatjie by “TLT’09” (Treebanks and Linguistic Theories) in Milaan, Italië. Dit is in die verrigtinge gepubliseer, met Jörg Tiedemann as mede-outeur. Die titel is “Building a Large Machine-Aligned Parallel Treebank”.
- Plakkaataanbieding by “TABU Dag 2009” in Groningen, getiteld “Training a statistical parser for parsing French for use in syntax-based machine translation”.
- Praatjie by “Computational Linguistics in the Netherlands (CLIN) 2009”. Titel: “Automatic filtering of parallel corpora for improving alignment accuracy”.
- Praatjie by ALASA-konferensie (International Conference of the African Language Association of Southern Africa) in Port Elizabeth, 9-11 Julie 2007. Titel: “Development of an Afrikaans wordnet: methodology and integration”.
- Praatjie by die gesamentlike konferensie van LSSA/SAALA/SAALT in Potchefstroom, 4-6 Julie 2007. Titel: “Ontwikkeling van ’n Afrikaanse woordnet: metodologie en integrasie”.

BEURSE, TOEKENNINGS, BEFONDSING EN ANDER

- Befondsing by die Universiteit van Suid-Afrika (Unisa) ontvang vir ’n postdoktorale beurs by die Akademie vir Afrikatale en Wetenskap (AALS) by Unisa, Pretoria (02/2014 tot en met 01/2016)
- Befondsing van die Nederlandse Taalunie ontvang vir deelname aan ’n masjienvertaalprojek, PaCo-MT, van 2008-2011 as deel van my PhD-program by die Rijksuniversiteit Groningen (2008-2013). Ná voltooiing van die projek in 2011 het die universiteit die res van my navorsingsaktiwiteite befonds tot aan die einde van 2012.
- ZASM-beurs (Zuid-Afrikaanse Spoorwegmaatschappij) vir Meestersgraadstudie in Amsterdam, Nederland (2005-2006)
- Rutgers-beurs van die Vrije Universiteit se Letterefakulteit vir Meestersgraadstudie in Amsterdam, Nederland (2005-2006)
- Projekbevondsing ontvang van die destydse Suid-Afrikaanse Departement van Wetenskap en Tegnologie vir die bou van ’n Afrikaanse woordnet (sien “OPLEIDING”) by die Sentrum vir Tekstegnologie (CTexT) by die Potchefstroomkampus van die Noordwes-Universiteit (2006)
- Lid van die “Landesstiftung Baden-Württemberg” in Duitsland (’n stigting wat geleerdes, studente en jong werkende mense in staat stel om in Baden-Württemberg, Duitsland te studeer) (2004)
- Gekies as uitruilstudent (verteenwoordigend van die destydse Universiteit van Port Elizabeth) vir die somersemester van 2004 by die Eberhard Karls-Universiteit Tübingen, Duitsland, met die gelyktydige toekenning van die Baden-Württemberg-beurs vir die uitruilprogram (2004)
- Nagraadse beurs by die destydse Universiteit van Port Elizabeth (2004)
- Dekaansbeurs by die destydse Universiteit van Port Elizabeth (2004)
- Geborg as deelnemer aan ’n Franse kursus en kulturele program by die Universiteit van Reunion-Eiland, Réunion (Frankryk) (2003)
- Lid van die “Golden Key International Honour”-genootskap (2003)
- DAAD-beurs (“Deutscher Akademischer Austausch Dienst”) vir Duitse studie in Suid-Afrika (2003)
- Merietebeurse by die destydse Universiteit van Port Elizabeth (2000-2003)
- Dux-leerling: Matriek 1999