

CURRICULUM VITAE

Gideon (Jozua) Kotzé, PhD

Current residence: Pretoria, South Africa

Email: kotzegi@ufs.ac.za

Website: www.gideonkotze.co.za

PERSONAL INFORMATION

Nationality: South African

Date of birth: 9 June 1981

Languages: In order of proficiency: Afrikaans (mother tongue), English, Dutch, German and French, limited knowledge of isiXhosa (high school education) and Japanese (spoken language course)

SUMMARY

I have a PhD in computational linguistics from the University of Groningen (2013), with 20 research publications. I have a passion for research in the digital development of language resources and technologies, especially with a focus on South Africa. The thesis describes my work on the automatic alignment of parallel syntactic trees, providing informed training data for syntax-based machine translation. I am currently developing a mobile app of place names with support for South African Sign Language at the University of the Free State, while also working on a web application for multilingual glossaries at the University of South Africa (Unisa). Research projects during my full-time research position at Unisa (2014-2019) include the development and semi-automatic creation of annotated XML corpora for South African languages, digitisation of physical texts, as well as statistical and neural machine translation. I was head of the Afrikaans wordnet project at the Centre for Text Technology (2006-7) and performed contract work for the same project in 2008. My Master's degree in Linguistics (Lexicology and Terminology) at the Free University Amsterdam (2005-6) also included an internship at the institution. At Unisa, I was also involved in the presentation of regular workshops on developing Wikipedia content for African languages. I have experience in writing grant proposals, supervising students, as well as teaching.

WORK EXPERIENCE

2021- Mobile app developer at the Department of South African Sign Language and Deaf Studies, University of the Free State, Bloemfontein, South Africa. Duties include the development of a cross-platform mobile application that accesses a database of South African place names in English and

South African Sign Language, as well as the design and implementation of said database. The project duration is from 1 December 2020 until 30 May 2022.

Research fellowship at the Department of South African Sign Language and Deaf Studies, University of the Free State, Bloemfontein, South Africa. Duties include research publications related to South African Sign Language and the obtainment of research grants. The fellowship is valid from 1 January 2021 until 31 December 2023.

Web application development at the Department of Geography at the University of South Africa, Pretoria. The project comprises a multilingual terminology database to be hosted by Unisa at the start of 2022.

2020 Project work for the South African Centre for Digital Language Resources (SADiLaR) through the Department of African Languages at the University of South Africa, Pretoria. Role involves automatic quality assurance of so-called wordnets for the African Wordnet project through the use of SQL database query and scripting.

Wrote an Afrikaans Wikipedia manual for the South African Academy for Science and Arts (Suid-Afrikaanse Akademie vir Wetenskap en Kuns) in Pretoria, together with Prof Laurette Pretorius.

2016-2019 Computational linguist (senior researcher) at the Academy of African Languages and Science (AALS), University of South Africa (Unisa), under Prof Laurette Pretorius, and, for the last few months until 30 June, Ms Nozibele Nomdebevana. Project activities include the implementation of a framework for TEI P5 corpus creation and quality assurance for source documents, the creation of a pipeline for the digitisation of Unisa institutional documents, machine translation experiments (statistical and neural), the latter of which involved isiZulu to English and Setswana to English, as well as collecting, collating, cleaning and filtering corpora and parallel corpora. During this time, as well as during the period of my postdoctoral fellowship, I was responsible for developing and updating the AALS website, and co-developed a website for the international 2018 Grammatical Framework summer school in Stellenbosch.¹ From 2015, I also contributed to and sometimes presented at monthly workshops that AALS held on campus where we assisted participants with contributing to Wikipedia in South African languages, and from 2014, I presented at various other occasions (see Talks and Presentations).

2014-2016 Postdoctoral Fellow at Unisa under Prof Laurette Pretorius. During this period, my research focused on statistical machine translation from isiZulu to English, as well as the design and implementation of a multilingual corpus building pipeline of Unisa institutional documents. In 2015, I

¹ <http://school.grammaticalframework.org/2018>

supervised two Honours students in a year module ([Honours Research Report - HRCOS82](#)) who completed their degrees in 2016 (BSc Honours in Computing and BCom Honours in Business Informatics).

- 2012 Designed and implemented a system using transformation-based learning to correct and apply high accuracy tree-to-tree alignments between phrase-structure trees. The goal is the creation of large-scale parallel treebanks for syntax-based machine translation, as with the PaCo-MT project (see below). This work was done as part of the research done for my PhD project (see Education).
- 2009 Taught a bachelor's level course at the University of Groningen "Tekstmanipulatie" (Text Manipulation) with Dr Barbara Plank.
- 2008-2011 I worked with a consortium on the machine translation project Parse and Corpus-Based Machine Translation (PaCo-MT) with the Catholic University of Leuven, Belgium, and the company *OneLiner bvba* in Sint Niklaas, Belgium. My main task consisted of the finding or development and subsequent application of tools to produce large parallel treebanks for training the system.
- 2008 Contract work completed for the Centre for Text Technology (CTexT), situated at the Potchefstroom campus of the North-West University in South Africa, with Prof Gerhard van Huyssteen as head of the centre. My focus was on the extension of the below-mentioned Afrikaans wordnet to a greater size, in which English entries were mostly translated into Afrikaans, while I applied scripting to perform data manipulation.
- 2006-2007 Temporary position as data developer and researcher at CTExT. Activities included: morphological analysis and quality control, the use of scripting to perform language data manipulation, the design and implementation of the *ALEXANDER* database (*Afrikaans Lexicon and Annotated Database for Engineering and Research*) and the Afrikaans wordnet, linguistic examination and quality control of lexica, corpus analysis and translation.
- 2006 Internship at the *Vrije Universiteit* (Free University) in Amsterdam as part of a master's degree. This involved the *Cornetto* project (*Combinatorial and Relational Network as Toolkit for Dutch Language Technology*), an electronic lexical database in Dutch for human language technological purposes. Various institutions were involved. Activities included semantic annotation and the use of scripting to perform text data manipulation.
- 2002-2004 Computer assistant/consultant at the Nelson Mandela University, Gqeberha, in computer labs for general use.

COMPUTER SKILLS

My background mostly involves experimental coding in research environments. Although light scripting is mostly sufficient, I also have experience in building more complex systems, such as the transformation-based learning tree aligner for my PhD and the mobile app at the University of the Free State.

I work mostly on Linux systems (Ubuntu, Mint, Fedora, Pop!_OS), MacOS (Lion up until El Capitan), as well as Windows. The currently running mobile app project at the University of the Free State involves development in a Windows 10 environment (Visual Studio 2019).

Programming

I program mostly in Python and also have some significant experience with Perl and Bash scripting. The current mobile app project at the University of the Free State involves C#, the Xamarin.Forms framework, as well as some Javascript and CSS, while the web application project at Unisa is built using the Django framework. In my university education I obtained some knowledge of C++, Java, and Turbo Pascal, while I have used some R for research purposes. I acquired some knowledge of XHTML, CSS, JavaScript, VBScript and ASP as part of a Web Technology course on Bachelor level. More recently, I also acquired more knowledge of HTML 5 and CSS 3 in a short online course, while implementing these technologies at work. I have experience with relational databases (Microsoft Access, MySQL, PostgreSQL, SQL Server). As a means of representing data I have much experience in the use of XML, XML parsers (Python's lxml.etree and Perl's XML::Twig) as well as some use of XPath, XSLT and JSON. Finally, I have experience using the version control software Git and use it on a daily basis.

Integrated development environments (IDEs) and/or IDE-like environments in which I have coded include Visual Studio, PyCharm, Eclipse, Atom, Emacs, as well as Sublime Text.

Computational linguistics tools

NLTK framework (sentence splitting, part-of-speech tagging, lemmatization, language recognition), the statistical machine translation toolkit Moses, the neural machine translation toolkit OpenNMT, more limited use of syntax-based machine translation (PaCoMT), word alignment tools (GIZA++, Fast Align), manual alignment tools (InterText, Stockholm TreeAligner), part-of-speech taggers (e.g. TreeTagger, Stanford, NCHLT²), syntactic parsers (Stanford, Berkeley, Alpino, MaltParser), the syntactic tree aligners Lingua-Align, Dublin Sub-Tree Aligner and TBLign (own software), other tools such as sentence splitters (Moses splitter, Punkt) and sentence aligners (Gale & Church, Hunalign), the document extraction, parsing and markup software GROBID (2018 workshop).

² Also tokenises, performs sentence boundary detection, and annotates with lemmas, named entities and phrase chunks.

Text production and relevant skills

Microsoft Office (Excel, Word, Access, Powerpoint), OpenOffice and LibreOffice (the text, spreadsheet and presentation features), as well as LaTeX for academic text production, using editors such as TeXShop and TexMaker.

Linguistic software

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

Other software

Image processing and Optical Character Recognition (OCR) tools: ImageMagick, Unpaper, Tesseract, Calamari, OCRopus, OCRFeeder, gscan2pdf. I have integrated all but the last two of these tools into a digitisation pipeline for processing institutional documents from Unisa (2016-2019).

EDUCATION

Institutions

- Rijksuniversiteit Groningen (University of Groningen) in Groningen, The Netherlands (2008-2012) (PhD)
- Vrije Universiteit (Free University) in Amsterdam, The Netherlands (2005-2007) (Master's degree)
- Nelson Mandela University (formerly the University of Port Elizabeth), Gqeberha (2000-2005) (B.A. and B.A. Honours)
- Eberhard Karls Tübingen University in Tübingen, Germany (2004) (student exchange programme)
- University of South Africa (General Linguistics I) (2001)

Degrees and courses

- 2011 Attended European Summer School in Logic, Language and Information (ESLLI) in Ljubljana, Slovenia.
- 2011 Attended a Machine Learning course at the University of Groningen, resulting in a paper with the title "Authorship attribution of weblog texts: A comparative study".
- 2010 Attended ESLLI summer school in Copenhagen, Denmark.
- 2010 Attended the 4th Machine Translation Marathon at Dublin City University in Dublin, 2010.
- 2009 Attended ESLLI summer school in Bordeaux, France.
- 2008-2013 **Doctor of Philosophy (Computational Linguistics)**
- 2013 Doctoral dissertation completed and accepted for public defense: *Complementary approaches to tree alignment: Combining statistical and rule-based methods*. In this work, the alignment of syntactic phrase-structures of translated texts is researched in order to create so-called *parallel treebanks*, of which the main application is syntax-based machine translation. Although the effect on the performance of such translation is found lacking, it is shown that various different novel approaches to the alignment problem can be successfully combined to create a high-quality resource.
- 2005-2007 **Master of Arts in Linguistics (Lexicology and Terminology)**
- 2007 Thesis completed: *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*. The work is a preliminary study about a proposed project at the North-West University in Potchefstroom, South Africa. It deals with the design and building of an Afrikaans wordnet – a semantic network stored as an electronic lexical database – for lexicographic and human language technological purposes. The project has been sponsored by the South African Department of Science and Technology.

- 2005-2006 Courses at the Vrije Universiteit (Free University) in Amsterdam, The Netherlands (names are in Dutch):
 “Lexicografische en Terminologische Tools” (Lexicographic and Terminological Tools), “Bilinguale Lexicografie” (Bilingual Lexicography), “Werkcollege Lexicologie” (Seminar in Lexicology), “Werkcollege Terminologie” (Seminar in Terminology), “Lexicale Semantiek” (Lexical Semantics), “Computationele Lexicologie” (Computational Lexicology) (no examination), “Relationele Modellen” (Relational Models), “Vaktaal” (Technical Language).
- 2004-2005 **B.A. Honours in Applied Language Studies (Translation Studies)** (*cum laude*)
- 2004 Courses at the University of Port Elizabeth:
 Introduction to General Linguistics, Applied Linguistics, Theory of Translation.
 Courses at the Eberhard Karls University in Tübingen, Germany:
 Computational Linguistics (Parsing and Programming in Java (Data Structures)), Advanced German, Spoken German.
- 2000-2003 **B.A. in Applied Language Studies** (*cum laude*)
- 2003 Courses: Afrikaans III (mainly Linguistics and Literature), German III, French III and Music Technology.
- 2001-2002 Courses: Afrikaans I and II, German I and II and French I and II.
 Additional Courses: Programming in C++, Programming in Java (Algorithmics and Data Structures, Web Technology, Mathematics (Algebra and Calculus), Information Systems and Computer Architecture.
- 2000 Courses: Afrikaans, Dutch, Introduction to University Practice and Computer Literacy.

ACADEMIC ACHIEVEMENTS AND EXPERIENCE

Publications

Accepted

Kapanadze, O., Kotzé, G., and Hanneforth, T. Accepted. Building Resources for Georgian Treebanking-based NLP. *Lecture Notes in Computer Science*. Springer.

2020

Kotzé, G. and Wolff, F. 2020. Exchanging image processing and OCR components in a Setswana digitisation pipeline. *South African Computer Journal*. Vol. 32(2), pp. 281-231. *South African Institute of Computer Scientists and Information Technologists (SAICSIT)*. doi: 10.18489/sacj.v32i2.707

2019

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

2017

Kotzé, G., Vandeghinste, V., Martens, S. and 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. and 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)*, pp. 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)*, pp. 48-53. IEEE. <http://ieeexplore.ieee.org/document/7813168/>.

2015

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

2014

Wolff, F. and Kotzé, G. 2014. Experiments with syllable-based Zulu-English machine translation. In Puttkammer, M. and Eiselen, R. (eds.): *Proceedings of the 2014 PRASA, RobMech and AfLaT International Joint Symposium*. pp. 217-222. Cape Town, South Africa.

Kotzé, G. and 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* (at LREC 2014), pp. 7-11, Reykjavík, Iceland.

2013

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

Vandeghinste, V., Martens, S., Kotzé, G., Tiedemann, J., Van den Bogaert, J., De Smet, K., Van Eynde, F., and Van Noord, G. 2013. Parse and Corpus-based Machine Translation. In *Peter Spyns and Jan Odijk (eds.): Essential Speech and Language Technology for Dutch*. pp. 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, pp. 71-96.

Kotzé, G., Vandeghinste, V., Martens, S. and 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)*, pp. 467-473, Istanbul, Turkey.

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, Latvia. NEALT Proceedings Series, Vol. 11, pp. 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*, pp. 122-127, Ljubljana, Slovenia.

Kotzé, G. 2011. Rule-induced error correction of aligned parallel treebanks. In *Proceedings of the International Conference "Corpus Linguistics - 2011"*, pp. 35-40, Saint Petersburg, Russia.

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

2009

Tiedemann, J. and Kotzé, G. 2009. A Discriminative Approach to Tree Alignment. In *Ilisei, I., Pekar, V. and Bernardini, S. (eds.): Proceedings of the International Workshop on Natural Language Processing Methods and Corpora in Translation, Lexicography and Language Learning* (in connection with RANLP'09), pp. 33-39. Borovets, Bulgaria.

Tiedemann, J. en Kotzé, G. 2009. Building a Large Machine-Aligned Parallel Treebank. In *Passarotti, M., Przepiórkowski, A., Raynaud, S. and Van Eynde, F. (eds.): Proceedings of the 8th International Workshop on Treebanks and Linguistic Theories (TLT'08)*, pp. 197-208. Milan, Italy.

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). pp. 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*. Master's Thesis. Free University of Amsterdam.

Accepted publications

Abstract accepted for presentation at the Southern African Linguistics and Applied Linguistics Society (SALALS) virtual conference: *SASL place-name planning: Developing an app as a non-official intervention*. 21-23 September 2021.

<https://salals.org.za/conferences/salals-virtual-conference-2021/>

Abstract accepted for presentation at the 2021 International Symposium on Place Names. *Finding place names: Improving digital documentation and accessibility of SASL place names*. (virtual) Bloemfontein. 29 September - 1 October 2021.

<https://www.ufs.ac.za/conferences/conference/2021ispn>.

Papers reviewed

- 1 paper for inclusion into the Proceedings of the 2017 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech)
- 1 paper for inclusion into the Proceedings of the 2019 SAUPEC / ROBMECH / PRASA Conference

Students supervised

In 2015 at Unisa, I supervised two Honours students in a year module (Honours Research Report - HRCOS82) who completed their degrees in 2016 (BSc Honours in Computing and BCom Honours in Business Informatics).

Courses presented

In 2009, I taught a Bachelor's level course at the University of Groningen "Tekstmanipulatie" (Text Manipulation) with Dr Barbara Plank.

Leadership positions

2014-2019 Academy of African Languages and Science, University of South Africa (Unisa) at the Muckleneuk Ridge main campus, Pretoria: leader of the digitisation and machine translation projects.

2006-2007 Centre for Text Technology (CTexT) at the Potchefstroom campus of the North-West University:

- Project leader of the lexical database *ALEXANDER* (Afrikaans Lexicon and Annotated Database for Engineering and Research) (2006)
- Project leader of the Afrikaans wordnet (2007 – see "EDUCATION")

TALKS AND PRESENTATIONS

- Presented two talks on using the Afrikaans Wikipedia for writing and editing at the South African Academy for Science and Arts (Suid-Afrikaanse Akademie vir Wetenskap en Kuns) in Pretoria, 30 January 2020.
- Presented preliminary findings of automatic quality assurance work for the African Wordnet project at a work session over two days at the Burgers Park hotel in Pretoria, 28 and 29 January 2020.
- Organised and presented a Wikipedia workshop on 3 April 2019 at the Potchefstroom campus of North-West University, who hosted the event, together with Dr Friedel Wolff and Ms Motswalle Kanyane, representing AALS, Unisa. The workshop was sponsored by the South African Centre for Digital Language Resources (SADiLaR).
- Oral presentation on 6 September 2018 at the DEASA 2018 conference at Unisa in Pretoria, entitled "A freely accessible, multilingual corpus for use in South African open learning environments".
- Oral presentation on 23 September 2016 at a Unisa seminar in Pretoria on the occasion of International Translation Day, with the title "The power of the text corpus: A Unisa perspective".
- Joint oral presentation with Friedel Wolff on 20 August 2015 at the National Conference on Multilingualism in Higher Education at Unisa in Pretoria, South Africa, with the title "Machine Translation at Unisa". Also co-authored a presentation at the same event titled "Building language resources from institutional content in Higher Education".
- Oral presentation on 26 June 2015 at the LSSA/SAALA/SAALT joint conference at the North-West University in Potchefstroom, South Africa, with the title "Applying English morphological segmentation to syllable-based Zulu-to-English statistical machine translation".
- Oral presentation on 22 April 2015 at Unisa's local "Machine Translation" seminar, organised by the Academy of African Languages and Science and the College of Graduate Studies. The title was "Rule-based syntactic tree alignment of parallel translated sentences", based on a selection of chapters from my PhD thesis.
- Oral presentation on 9 July 2014 at Unisa's local "Sustainable Multilingualism Seminar", organised by the Academy of African Languages and Science and the College of Graduate Studies. The title was "Experiments with syllable-based English-Zulu alignment", based on one of my papers with the same title.
- Oral presentation on 27 May 2014 at the "SaLTMiL Workshop on free/open-source language resources for the machine translation of less-resourced languages" at the Language Resources and Evaluation conference (LREC 2014) in Reykjavík, Iceland.
- Defence of doctoral dissertation at the University of Groningen on 24 June 2013, resulting in the conferment of a PhD degree.

- Poster presentation at the Student Session of the European Summer School in Logic, Language and Information (ESLLI 2011) in Ljubljana, Slovenia. The title is "Improving syntactic tree alignment through rule-based error correction". (published)
- Oral presentation at Corpus Linguistics 2011 in Saint Petersburg, Russia. The title is "Rule-induced error correction of aligned parallel treebanks." (published)
- Poster presentation at the Nordic Conference on Computational Linguistics (NoDaLiDa'19) in Riga, Latvia, on 11-13 May 2011. The title is "Finding Statistically Motivated Features Influencing Subtree Alignment Performance". (published)
- Talk at the Computational Linguistics in the Netherlands (CLIN 2011) conference in Ghent, Belgium. The title is the same as the above: "Finding Statistically Motivated Features Influencing Subtree Alignment Performance".
- Invited in 2010 by Prof. Dr. Martin Volk as a colloquium speaker at the Institute for Computational Linguistics at the University of Zürich. The topic was the tree-to-tree alignment tool Lingua-Align developed by Jörg Tiedemann. Here, I presented the published conference paper authored and presented by Tiedemann at LREC 2010 in Malta.
- Oral presentation at TABU Dag 2010 in Groningen. Title: "Supervised tree alignment for syntax-based machine translation".
- Poster presentation at Computational Linguistics in the Netherlands (CLIN) 2010 in Utrecht, The Netherlands. Title: "Discriminative parallel treebank alignment for syntax-based machine translation".
- Oral presentation presented at TLT'09 in Milan, Italy. Published in proceedings and co-authored with Jörg Tiedemann. Title: "Building a Large Machine-Aligned Parallel Treebank".
- Poster presentation at TABU Dag 2009 in Groningen. Title: "Training a statistical parser for parsing French for use in syntax-based machine translation".
- Oral presentation presented at Computational Linguistics in the Netherlands (CLIN) 2009. Title: "Automatic filtering of parallel corpora for improving alignment accuracy".
- Oral presentation presented at ALASA conference (International Conference of the African Language Association of Southern Africa) in Port Elizabeth, South Africa, 9-11 July 2007. Title: "Development of an Afrikaans wordnet: methodology and integration".
- Oral presentation at LSSA/SAALA/SAALT joint conference (Linguistic Society of Southern Africa (LSSA), the South African Applied Linguistics Association (SAALA) and the South African Association for Language Teaching (SAALT)) in Potchefstroom, South Africa, 4-6 July 2007. Title (in Afrikaans): "Ontwikkeling van 'n Afrikaanse woordnet: metodologie en integrasie" (Development of an Afrikaans wordnet: methodology and integration).

SCHOLARSHIPS, AWARDS, FUNDING AND OTHERS

- Interdisciplinary grant awarded from the University of the Free State to research team at the Department of South African Sign Language and Deaf Studies (2021).
- Research fellowship awarded from the University of the Free State, Bloemfontein, South Africa, for research performed at the Department of South African Sign Language and Deaf Studies, (01/01/2021 until 31/12/2023).
- Postdoctoral fellowship awarded at the University of South Africa (Unisa), Pretoria, for research performed at the former Academy of African Languages and Science, College of Graduate Studies (02/2014 until 01/2016)
- Obtained funding from the Dutch Language Union for partaking in a machine translation project, PaCo-MT, from 2008-2011 as part of my PhD program at the University of Groningen (2008-2013). After completion of the project in 2011, the University funded the rest of my research activities up to the end of 2012.
- ZASM (Zuid-Afrikaanse Spoorwegmaatschappij) scholarship for Master's study in Amsterdam, The Netherlands (2005-2006)
- Rutgers scholarship from the Vrije Universiteit's Department of Arts (Letteren) for Master's study in Amsterdam, The Netherlands (2005-2006)
- Project funds received at the South African Department of Science and Technology for the construction of an Afrikaans wordnet (see "EDUCATION") at the Centre for Text Technology (CTexT) at the Potchefstroom campus of the North-West University, South Africa (2006)
- Member of the *Landesstiftung Baden-Württemberg* in Germany (a foundation enabling scholars, students and young working people to study in Baden-Württemberg, Germany) (2004)
- Chosen as exchange student (representing the former University of Port Elizabeth) for the summer semester of 2004 at the Eberhard Karls University of Tübingen in Tübingen, Germany, with the simultaneous conferment of the Baden-Württemberg scholarship for the exchange programme (2004)
- Postgraduate scholarship at the University of Port Elizabeth (2004)
- Dean's scholarship at the University of Port Elizabeth (2004)
- Sponsored as participant in a French course and cultural programme at the University of Reunion Island, Réunion (France) (2003)
- Member of the Golden Key International Honour Society (2003)
- DAAD (*Deutscher Akademischer Austausch Dienst*) scholarship for German study in South Africa (2003)
- Merit bursaries at the University of Port Elizabeth (2000-2003)
- Dux pupil: Matric 1999 (best position in final year at high school)