

Swearing in South Africa: Multidisciplinary research on language taboos

Van Huyssteen, Gerhard B
Centre for Text Technology (CTeX:T), North-West
University, Potchefstroom, South Africa
gerhard.vanhuyssteen@nwu.ac.za

Abstract

Research on swearing (used here as a hypernym to include other phenomena and/or synonyms like cursing, profanity, taboo language, etc.) has been prevalent for many years internationally, also from a variety of scientific disciplines. Most of the research literature, however, is on swearing in English, although studies have also been conducted on some other languages. By contrast, very little to no research has been done on swearing within the South African context, which is quite surprising, given that using certain swearwords (i.e., racial slurs) is punishable by law.

To address this void, we established a multidisciplinary project with its primary roots in the digital humanities, and with inputs from and implications to (amongst others) linguistics, literary studies, communication studies, neurology, psychology, sociology, computer sciences, and law. This project (and specifically the topic of swearing) holds the potential to provide insights in human cognition and social interaction, while situating it broadly within the scope of the Fourth Industrial Revolution. The project commenced in July 2019, and is currently ongoing.

In this paper, we firstly provide a rationale for the project, before introducing each of the five subprojects. These subprojects pertain to swearing and the law; a swearing constructicon (a kind of online dictionary) for Afrikaans; swearing in the entertainment world and in the media; swearing as a linguistic innovation; and an end-user facing project website. We also report on some of the outputs from the project that are already available, and others that are still being developed and investigated. We conclude with a brief overview of some of the potential impacts of the project.

Keywords: censorship, computational linguistics, cursing, language change, taboo

1 Introduction

Swearing is a fascinating phenomenon that not only gives us deep insights in human cognition and neurophysiology, but also in social interactions and power dynamics. However, very little multidisciplinary research has been done on swearing in the South African context – a lacuna that the project *What the Swearword?! (WTS)* aims to fill with insights from the digital humanities, and with inputs from and implications to linguistics, literary studies, communication studies, psychology, neurology, sociology, computer sciences, and law. The project commenced in July 2019 with a three-year set-up and exploratory phase (focusing only on Afrikaans, and other languages in its ecosystem – including other Germanic languages), ending in June 2022. Thereafter, the project will continue in directions determined by the interests of the multidisciplinary team members, and depending on the availability of funding.

The following types of (popular) questions are of interest to researchers in the project:

- If a website contains swearing, what legal obligations does the owner/developer have?
- Should parents protect their children from hearing swear words?
- What is the best way to determine objective offensiveness ratings for swearwords, e.g., to determine advisories for films and/or books?
- How does it happen that an Afrikaans word like *be:fok* (a verbalized form of *fuck*) can mean, among others, both ‘good’ (as in *Dit was nou befok genees!* ‘That was really fucking A’), and ‘angry’ (as in *Hy is al weer befok!* ‘He is once again fucked off!)?
- How is swearing used as a linguistic innovation that causes short-term and/or rapid language change?
- What are the views on swearing of writers, dramatists, poets, TV and film makers,



producers, directors, actors, musicians, editors, journalists, podcasters, bloggers?

- How and why do these content creators apply self-censorship with regards to swearing? What is the impact of cancel culture on their language usage in the content they create?
- What is the interaction between swearing and societal change?
- What is the neurological impact when someone hears a racial, homophobic, or sexist slur?

In addition to the primary focus on swearing, the project also has a secondary, subjacent aim, namely, to investigate alternative, contemporary opportunities of scholarly communication, specifically focusing on podcasts, blogs, videos, and webinars. Traditional main-stream outlets for communicating research results, i.e., monographs, edited books, journal articles, conference publications, and presented talks and posters, are by and large still the only research outputs that carry weight in academic appointments and promotions, and in the national and international evaluations of universities. This is especially true for the humanities and social sciences, and even more so in the South African context. A fundamental (albeit radical) presupposition of this project is that these main-stream outlets for communicating research results are already outdated and will become even more outdated and less appropriate in a technologically revolutionized society [1]. We therefore aim to experiment with how to incorporate and integrate peer-reviewing in new communication channels (to ensure quality); how to optimize such means to stimulate multidisciplinary interest and foster new collaborations; and how to use these channels to enable and fast-track research (e.g., increasing respondent participation).

The aim of this paper is twofold: (1) To provide rationales for each of the subprojects; and (2) To report on some of the outputs and milestones of the project after two years of research and development. The overarching theme is that the digital humanities afford one with even more opportunities to stimulate multidisciplinary in and outside the humanities. In the next section, we

give a brief overview of previous research on the topic, indicating that there is a lacuna in knowledge on, and understanding of swearing in the South African context. In Section 3, each of the five subprojects are introduced, while we report on some of the outputs in these subprojects in Section 4. We conclude with a brief perspective on some of the other benefits and impact of the project.

2 Multidisciplinary research on swearing

For many decades, swearing (used here as a hypernym to include other phenomena and/or synonyms like cursing/cussing, profanity, blasphemy, obscenity, vulgarity, verbal abuse, verbal sparring, (racial) slurs, terms of abuse, insults, dirty language, and taboo language) has been researched internationally from various disciplines, including literary studies, journalism and communication studies, psychology, sociology, law, philosophy and ethics, cultural anthropology and history, pediatrics, neurology and other neurosciences. In linguistics specifically, studies range from comparative etymology, lexicology and lexicography, typology, and grammar, to first- and second-language acquisition, variation studies and dialectology, and sign-language, gestures and kinesics. Interdisciplinary research is often conducted within the fields of sociolinguistics, psycholinguistics, computational linguistics, and neurolinguistics. It is true that most of the literature is on swearing in English, although studies have also been conducted on many other languages, such as Cantonese, Danish, Dutch, Finnish, French, Italian, Japanese, Latin, and Russian, amongst many others. The titles of a few seminal and/or recent books serve to illustrate: *The Oxford Handbook of Taboo Words and Language* (Allan 2019); *Advances in swearing research: New languages and new contexts* (Beers Fägersten & Stapleton 2017); *What the F – What swearing reveals about our language, our brains, and ourselves* (Bergen 2016); *Why we curse: A neuro-psycho-social theory of speech* (Jay 2000); *Nine nasty words: English in the gutter: then, now, and forever* (McWhorter 2021); *Offensive Language: Taboo, offence and social control* (O'Driscoll 2020); *Linguistic Taboo Revisited: Novel Insights from Cognitive Perspectives* (Pizarro Pedraza



2018); and *Rot lekker zelf op: Over politiek incorrect en ander ongepast taalgebruik* (Van Sterkenburg 2019).

By contrast, very little to no research has been done on swearing within the South African context, which is quite surprising, given that using certain swearwords (i.e., racial slurs) is punishable by law. Most of the linguistic research has focused on the lexicographic treatment of swearing (e.g., Dekker 1991; Van Huyssteen 1998), while only a handful of studies focused on grammatical aspects of swearing (e.g., Calitz 1979; Feinauer 1981; Van Huyssteen 1996). Most recently, Van der Walt's (2019) MA dissertation at the North-West University (NWU), had a section on swearing as part of her analysis of Zefrikaans (an informal variety of Afrikaans). In other fields, research has also been sparse; for example, in Coetzee's 2018 article on children's swearing in multilingual contexts, there are only three references to other (socio)linguistic research that has been conducted in the South African context.

To address the lacuna in knowledge on, and understanding of swearing in the South African context, we conceptualized five initial subprojects; the rationale for these is discussed in the next section.

3 Subprojects

3.1 A: Swearing and the law

The South African Film and Publication Board (FPB) regulates age restrictions on films, computer games, and publications that don't fall under the jurisdiction of the Press Ombudsman, which are released/published in South Africa. One of their criteria relates to what they call "strong language", which is defined as "crude words, threats, abuse, profanity or language that amounts to prejudice" (Republic of South Africa 2019). They will add the label "L" to a film, computer game or publication to alert users that there is use of strong language "of a mild, moderate, strong or very strong impact". However, this offensiveness scale is nowhere operationalized.

Following from this, several questions arise (to mention but a few):

- Can these categories of the FPB be predicted automatically (e.g., through machine learning algorithms)?
- Should adults and children be treated differently regarding swearing? Is swearing considered "adult/mature content", or simply as "explicit content"?
- These guidelines refer specifically to films and computer games, but what about other media, such as websites, literary texts, memes, songs/lyrics, and podcasts with swear words? Should these also carry content advisories? What are end-users' (e.g., parents) expectations about such advisories?
- What about swear words/text linked to images, videos and/or sound? For example, what about swearwords in lyrics and music videos?
- Given the history of censorship in South Africa (Van Rooyen 2012), how should we balance freedom of speech and freedom of choice, vs. protecting the citizens (e.g., children) of South Africa?

3.2 B: *Vloekopedia*: An encyclopaedic constructicon of Afrikaans swearing

Dictionaries and encyclopedias of swearing in English, Dutch, Spanish, Cantonese, Russian, etc. abound, none exists for Afrikaans and/or other indigenous South African languages. In addition, many of the dictionaries and encyclopedias for other languages are not authoritative, but mainly presented as popular entertainment (with two notable exceptions: Hughes (1991), and Sheidlower (2009), with the latter restricted to only the word *fuck* and its compounds and derivations). To address this lacuna for Afrikaans, we commenced to compile an encyclopaedic constructicon of Afrikaans swearing, called *Vloekopedia*.

Theoretically, the *Vloekopedia* will be underpinned by cognitive construction grammar, specifically as a constructicon, which is "a theoretical conception of language as a structured inventory of constructions, and ... a collection of construction descriptions, essentially a practical instantiation of the former concept" (Lyngfelt *et al.* 2018:1). The



idea of a construction as a “dictionary of constructions” was first suggested by Fillmore *et al.* (2008), which subsequently led to construction projects for Brazilian Portuguese, German, Japanese, Russian, and Swedish. Constructicography is a blend between construction grammar and lexicography, with the aim to compile a practically usable descriptive resource of lexical, morphological and/or syntactic constructions.

A central tenet of cognitive construction grammar is that it is usage-based, i.e., the view that constructions are generalisations over specific, real-world instances, based on, among others, frequency and salience. To identify and describe constructions, methods from corpus linguistic and/or psycholinguistic are most often used; this approach is therefore in its very essence suitable for multidisciplinary research.

Another important principle of cognitive construction grammar is its view of semantics being encyclopaedic, i.e., that meaning cannot be captured by means of a (lexical) definition only. Instead, usage patterns, pragmatics, associations, inferred knowledge, cultural importance, etc. are all part of the conceptual “meaning” of words and expressions. It is admittedly difficult (if not impossible) to capture such vast knowledge of constructions in the form of a (linear, linguistic) dictionary, but one could at least attempt to include elements such as real-world examples, frequency-based collocations, extensive pragmatic tags, mixed media, related information from other languages in the ecosystem, etymological information, etc.

One particular type of encyclopedic information that we are focusing on (also in relation to subproject A), is the rating of swear words and expressions on a taboo scale. To obtain offensiveness ratings for words has been done for a few languages (see Beers Fägersten (2007; 2012) for an overview), but never before for Afrikaans.

3.3 C: Swearing in the entertainment world and media

One of the landmark cases in censorship in the South African context, was the banning of *Magersfontein, O Magersfontein!* (Leroux 1976) in

1977. The main arguments for banning the book were based on the language in the book: “... excessive foul language, excessive vain use of the Name of the Lord, vulgar references to defecation, masturbation, loss of virginity, prevention of conception by rinsing with soap, menstruation, genitals and prostate trouble ...” [translated] (Leroux 1990). Much has changed since the fierce grip that the Film and Publication Board had on South African entertainment and media in the 1970s and 1980s.

To our knowledge, no focused research has ever been done (also not recently) on why and how content creators use swearing in entertainment and the media. In this subproject, we therefore investigate the views on swearing of content creators in entertainment and the media (e.g., writers, dramatists, poets, TV and film makers, producers, directors, actors, musicians, editors, journalists, podcasters, bloggers). Of special interest, is how they are potentially impacted by the current cancel culture (as a form of social censorship).

3.4 D: Swearing, linguistic innovation, constructionalisation, and language change

Linguistic innovation (a.k.a. linguistic creativity) as an instigator of language change has been studied widely in linguistics. With regard to linguistic innovation in the 21st century, Paradowski & Jonak (2012) note that “[e]rstwhile research on language evolution and change focused on large timescales, typically spanning at least several decades. Nowadays, observable changes are taking place much faster. According to the Global Language Monitor (2009) a new English word is born roughly every 98 minutes ...” Analyses of linguistic data from so-called Web 2.0 sources (e.g. blogs, microblogs, social media, and comments on websites) potentially provide us with insight into complex, dynamic systems, including “society, variations and typology, the rise of new grammatical constructions, semantic bleaching, language evolution in general, and the spread and competition of both individual expressions, and entire languages ...” (Paradowski & Jonak 2012).

For example, in a post on Facebook on 7 April 2019 the user Don Dapper commented on a



Figure 1: The "what in the X fuck" construction



fashion photo of a person wearing accordion-like attire: “What in the accordion FUCK is this???” (See Figure 1). Two days later, someone in a WhatsApp group commented on a picture of the elevation profile of a half-marathon (see Figure 1): “What in the steep cliff FUCK are they talking about!!”. The first expression could be considered a syntactic extension of the expression *what the fuck*, which in itself could be considered a syntactic innovation – i.e., *what the X* is only used in contexts where X could be filled with a swear word (or euphemism). Similar swearing-specific constructions could be observed in Afrikaans (e.g., *de X in*, as in *de moer in*, *de bliksem in*, *de fok in*, etc.), Dutch (e.g. *krijg de X*, as in *krijg de tyfus*, *krijg de rambam*, *krijg de pokken*, etc.), or English (e.g. *by X!*, as in *by God!*, extended to *by Toutatis!* or *by Jupiter!* in the Asterix comic book series).

This subproject has the strongest linguistic focus of all subprojects, since we investigate morphological and syntactic constructions that are specific to the domain of swearing, i.e., part of a swearing constructicon. One of our main interests is how new constructions are continuously added to the constructicon via the process of constructionalisation (Traugott & Trousdale 2013). In this regard we also focus on the role of cross-linguistic constructionalisation (Höder 2018), specifically focusing on Afrikaans, English, Dutch (and potentially French with relation to Flemish Dutch). For example, is the above-mentioned *de X in* construction the source for Afrikaans *wat de fok!*, or is it rather the result of transfer from English *what the fuck!*? Or is it a

combination of both? This part of the research is not only relevant to the swearing domain, but also more generally to language change (in contact situations).

The importance of traditional social networks as a determining factor in language change has been accepted widely in linguistics (see Labov (2001), for instance). In recent years, the role of modern social networks (in the form of social media) has gained prominence in research on rapid linguistic change (e.g., Goel *et al.* 2016). The basic idea is that linguistic innovations can potentially gain momentum in speech communities more rapidly and widespread through social media, than is the case in traditional social networks and through traditional media. By analyzing unedited linguistic data from social media (e.g., Twitter, Facebook, Reddit, or comments on blogs, newspaper articles, etc.), we can therefore potentially observe language change “as it happens”.

3.5 E: *Vloekcoza*: project website and social media presence

It is not uncommon for research projects to have independent websites with unique, easy-to-remember URLs. We have therefore set up a secure, technology-rich, end-user facing project website, *vloek.co.za*, as a means to create awareness of and cultivate new collaborations on the project, to publish outputs from the project, and to create a platform where registered users can participate in the above-mentioned surveys. In order to create a wide awareness of the website, we have also created project pages on Facebook, Twitter, Instagram, and Pinterest. Our main focus, however, is on Facebook, where an additional group, *Vloek*, has been established. This group serves as the first stop to gather information and data from end-users, as well as to disseminate information.

4 Progress and outputs

4.1 Subproject A (law)

Since the one of overarching questions of this subproject is how to classify Afrikaans swearwords according to the categories identified by the South African FPB, the main output of this subproject is the *Vloekmeter* (‘swearing meter’; see



Figure 2: *Vloekmeter* showing results for "fokken" ('fucking') and "frieken" ('fricking')



vloek.co.za/vloekmeter). The *Vloekmeter* is purely data-driven: Based on data from single word surveys (SWSs), statistics are presented on an interactive dashboard on the website (see Figure 2). In each SWS, only one swearword is presented to registered participants. The aim with SWSs is to keep each one as short as possible, in order to prevent respondent fatigue (Lavrakas 2008). The assumption is that one would cover more words over a period of time, than if one were to present the same number of words to participants in a single session. As of 15 August 2021, 51 such SWSs have been posted, with a total of 6 243 responses (an average of 122.4 responses per SWS). These results have already been used for research on statistic modelling in the digital humanities ([REFERENCE 1 REMOVED]), as well as lexicology studies ([REFERENCE 2 REMOVED]).

This subproject also provided the impetus for two master's degree students currently working on their dissertations. One of the students (Mart-Mari van der Merwe; University of Pretoria (UP)) is identifying the 50 most prototypical Afrikaans swearwords, in order to obtain offensiveness ratings / taboo values for them. Another student (Colette Combrink; (NWU)) focuses on cancel culture as a form of social censorship, and how it impacts on a variety of writers and authors.

At the beginning of 2021, we (in collaboration with Maroela Media, the largest Afrikaans online news publication, and WatKykJy, a very free-thinking Afrikaans blog site) have conducted a large-scale survey to determine what the attitudes of adult Afrikaans speakers are towards content advisories for films and books (e.g., indication of suitability for certain age groups, themes covered, etc.). These results are currently (as of 15 August 2021) being processed and interpreted, and will be published during 2021/2.

4.2 Subproject B (*Vloekopedia*)

Until now, this subproject has focused only on data collection, and more specifically on lexical items (i.e., words, rather than phrases and expressions). A core lexicon of 711 words has been compiled in 2019/20, mainly based on data from WatKykJy. It was supplemented with data crawled from UrbanDictionary, resulting in 131 additional usable entries. Both datasets were manually curated by a student assistant.

In 2021, the *Woordeboek van die Afrikaanse Taal* (WAT), *Handwoordeboek van die Afrikaanse Taal* (HAT) and Centre for Text Technology (CTeX) of the NWU agreed generously to supply the project with relevant material from their respective databases. This data was amalgamated with the above-mentioned data, to construct a single

database consisting of 3,858 entries (as of 15 August 2021). Subsequently, one of the computational linguists on the project (Jaco du Toit) wrote a complex script to not only retrieve frequencies for all entries from all available corpora on VivA's Corpus Portal (VivA 2021), but also to extract all examples where these entries occur. This resulted in a database of 273 MB, containing more than 3,5 million sentences. This database needs to be curated, which in itself will be a gigantic task. Currently a master's student (Mart-Mari van der Merwe) is working on solutions to clean-up at least a portion of the data. Work will continue into the foreseeable future.

4.3 Subproject C (entertainment/media)

In addition to work already mentioned under subproject A, work in this subproject has focused by and large on the production and release of the podcast series *Wat de Vloekwoord?!*. This is a podcast series that explores the views and attitudes of content creators in the entertainment world and media on swearwords and taboo topics. Through interviews with well-known (Afrikaans) writers, TV and filmmakers, directors, actors, musicians, editors, journalists, podcasters and bloggers, we explore censorship in South Africa, what the function of swearing is, how viewers and listeners respond to swearwords, and so on.

The first episode of the first season was launched on 4 September 2020; the fourteenth (and last) episode of the first season was published on 18 December 2020. The series was co-hosted by psychologist me and Elmarie Claassens (clinical psychologist), and was technically produced by Gifford Peché (Decibel Studios). The first season consisted of interviews with prominent figures in the South African entertainment and media industry, including Anton Goosen, Amanda Strydom, Claire Johnson, Neil Sandilands, and Hunter Kennedy (to name but a few).

On Anchor.fm (the platform where the podcast is hosted), these fourteen episodes have been played a total number of 2,704 times (an average of 195 times per episode, as of 15 August 2021). Planning for a second season of fourteen episodes in a different format has commenced. The second season should launch in September 2021.

In collaboration with Afrikaans.com, a campaign related to this project was run from August till November 2020. This project not only created awareness of the project (with a significant increase in the number of registered users), but also promoted three questionnaires related to swearing in entertainment and the media. In addition, five blogs by renowned journalists (and one student) have been published on vloek.co.za. All of these blogs centered around the theme of swearing in the media, including Afrikaans music, radio, and newspapers.

4.4 Subproject D (linguistics)

Being one of the central subprojects of this project (since the main project's focus is on a phenomenon that manifests in language usage), and since many of the members are trained and/or practicing linguists, it is expected that this subproject will be the long-term focus of the main project. Hence, this is also the subproject where most of the fundamental "thinking" about directions for the other subprojects happens. Despite its central role, it is however the subproject with the least number of outputs to date, but the one with the most important outputs (in my personal opinion).

The first important output that this subproject directly lead to, is the establishment of an honors-level course in linguistics, called *Pornolinguistics: Swearing and other language taboos in cognitive neurosciences*. This course was conceptualized as a collaboration initiative between the departments of Afrikaans (and Dutch) at NWU and UP, re-utilizing existing course modules at both institutions. Virtual teaching and learning – due to the Covid-19 pandemic – played a central role in establishing the course, since it became more "natural" for students from two universities to be in the same virtual classroom, while it also afforded the opportunity to involve many other experts to teach specialized sections of the course.

The course was designed around four disciplines, with specific themes in each of these (see Table 1). Aside from lecturers from the above-mentioned departments of Afrikaans (and Dutch), lecturers also included computational linguists from the NWU's CTeX; a pediatric neurologist,



Table 1: Honors module

Discipline	Theme
Introduction	<ul style="list-style-type: none"> • What is swearing and language taboos? • What is cognitive neurosciences?
Linguistics	<ul style="list-style-type: none"> • The constructicon and constructicography • Construction grammar • Constructionalization and subjectification • Methodology: Sociolinguistics • Methodology: Corpus linguistics
Computer sciences	<ul style="list-style-type: none"> • Artificial intelligence • Sentiment analysis • Hate speech recognition
Neurology	<ul style="list-style-type: none"> • Neuroanatomy and language • Coprolalia and other disorders • Neuro-imaging
Psychology	<ul style="list-style-type: none"> • Emotion • Language acquisition

psychologist, and speech therapist from NWU's Centre for Health and Human Performance (CHHP); and a clinical psychologist in private practice.

Five students enrolled for the course in 2021, while a number of guests also joined the classes per occasion. As part of the course outcomes, students have to write blogs, popular articles, a research proposal, and a conference presentation (among others). Many of these outputs will be submitted for publication towards the end of 2021.

The course will continue in 2022, but with two additional opportunities:

- (1) All lectures will be presented as public symposia, in order to enable external people to also attend these lectures.

- (2) The linguistics section will also be attended by students from the University of Leiden (The Netherlands) who are enrolled for a postgraduate course in Afrikaans linguistics.

The second output of this subproject, is the establishment of a research discussion group on construction grammar and constructionalization, consisting of linguistics researchers from NWU and UP. Members of the group meet virtually for a weekly discussion session of 90 minutes, where they attend online courses together, discuss recent publication, and work together on research outputs. Two presentations at international symposia during October 2021 have already been accepted, while the first scholarly publications from this group is scheduled for submission before the end of 2021.

4.5 Subproject E (Vloekcoza)

Setting-up, designing, developing and implementing the project website and associated social media pages, took up most of the financial and other resources during the first twelve to eighteen months of the project. Since the project website is meant to be a fully functional, secure, technology-rich, end-user facing product, it was of utmost importance to ensure that it is a secure platform, is able to handle traffic, can accommodate various kinds of posts, is easy to maintain by non-technical people, can work well on mobile devices, etc.

In addition to the main functionalities of the website, a complete end-user facing, online booking system for the podcast series have been developed by BlueTek Computers. This system enables the interviewers to interact in a professional and systematic way with podcast guests, specifically to make bookings for online or personal interviews, obtain official permission for release of podcasts from guests, etc.

In our assessment, the initial investment in this subproject was well-worth the time and money. One of the best dividends is that the project has a dedicated platform to host a variety of surveys for data collection; as was mentioned in 4.1, we have already been able to publish more than 50 surveys, with more than 120 responses on average per survey (also see Table 2). In some disciplines this



Table 2: End-user interactions of *Vloekcoza*

Interactions	2019/20	% increase	2020/21	% increase
Registered users on website	1 434	100%	2 075	31%
Questionnaires	30	100%	51	41%
Responses to questionnaires	4 801	100%	6 243	23%
Facebook group: Members	553	100%	708	22%
Facebook page: Likes	264	100%	359	27%
Facebook page: Followers	281	100%	385	27%
Instagram: Followers	158	100%	218	28%
Twitter: Followers	13	100%	21	33%
Pinterest: Followers	4	100%	25	84%
Pinterest: Engaged audience	0	0%	659	100%
Anchor.fm: Plays	0	0%	2 704	100%

might be deemed a small response rate, but for the kind of data we are collecting, it is quite substantial. For the sake of comparison, for similar research conducted by Beers Fägersten (2012), she only used 60 respondents from one university campus.

Table 2 also presents some of the other interactions with end-users. The percentage of increase for the first period (September 2019 to August 2020) is a 100% in all cases, since the project started with zero interactions. A slow, but steady growth can be observed for the second period of reporting (September 2020 to August 2021). We are confident that this trend will continue in the years to come.

5 Conclusion

In addition to the above tangible outputs, the project also have (potential) impact in other ways:

- In addition to creating opportunities for post-graduate students, the project has also created part-time job opportunities for student assistants (one per year), and a web and social media editor (one per year).
- Since one of our secondary aims is to foster collaboration outside our “usual” disciplines and networks, the project has already shown its potential to create such new opportunities. We hope that this will increase substantially in future, with collaboration with even more disciplines, more institutions, and more countries.

- All the data and corpora that have been, are being, and will be developed during the course of this project, will be made available for distribution under an open-source license by the South African Centre for Digital Language Resources (SADiLaR), so that it could eventually be utilized in many other follow-up or competing projects.
- Given the priority of the development and integration of new and emerging indigenous ICTs, as well as an exponential rise in interest in artificial intelligence research and development, this project stimulates the conceptualization, design, development and implementation of new resources and technologies (at least for Afrikaans, until now). We believe that it holds the potential to also attract, expand, and support research in the digital humanities as part of the process of building South Africa's information society.

Notes

[1] To illustrate this presupposition with two examples: (1) South African’s National Skills Fund (NSF) CEO Mvuyisi Macikama said in October 2018 that the NSF has a target of training 30,000 artisans a year by 2030, and that students in the social sciences and humanities do not contribute to the job market. He argued that funding should therefore be channeled away from “soft degrees” offered by universities (MyBroadband 2018). Although we do not subscribe to this view in any



possible way, it does illustrate the perception that social sciences and humanities are irrelevant in the South African context. Perhaps if research results from these disciplines have been more visible (and “digestible”), perceptions like these could be changed over time.

(2) In his now widely known (albeit controversial) article, Meho (2007) stated: “It is a sobering fact that some 90% of papers that have been published in academic journals are never cited. Indeed, as many as 50% of papers are never read by anyone other than their authors, referees and journal editors.” These statistics have been challenged and re-evaluated by numerous other scholars, but Remler (2014) still concludes that “[a]cademic publishing needs fixing”, especially since more than 80% of articles in humanities are never cited. This trend to “fix” academic publishing is seen in numerous other forms, including a strong drive towards open-access publication. For example, in March 2019 the University of California cancelled its subscription to Elsevier, the world’s largest publisher of academic journals, as part of their crusade to transform scholarly communication.

If one looks more closely at linguistics, and specifically at linguistics in (South) Africa, one could for example note that on the renowned Scimago Lab’s list of journals (www.scimagojr.com), only six linguistics journals from the continent appear with a Scimago Journal Rank (SJR) indicator, with *Lexikos* rated highest (SJR=0.280 for 2020), and *Stellenbosch Papers in Linguistics Plus* lowest (SJR=0.104 for 2020). To take the latter for illustration purposes, it means that 48 articles were published between 2017-2019, and these articles were only cited 8 times up to 2020. One of the most renowned international linguistics journals, *Language*, published 139 articles between 2017-2019, and these were cited only 240 times up to 2020.

In all honesty, one should take cognizance of the fact that, in contrast to publications in the natural sciences with relatively quick turn-over times, humanities journal articles are typically cited over a longer period of time. In addition, citations in books are often counted to a limited degree (or not at all), and this potentially has some effect on impact and evaluations. Nonetheless, these quoted

figures illustrate a general tendency regarding publications in linguistics and the humanities specifically.

Acknowledgements

This research is partially funded by the Suid-Afrikaans Akademie vir Wetenskap en Kuns, and partially made possible through barter agreements with BlueTek Computers, Afrikaans.com, and WatKykJy.co.za. The *Woordeboek van die Afrikaanse Taal* (WAT), *Handwoordeboek van die Afrikaanse Taal* (HAT) and Centre for Text Technology (CTeX) of the North-West University are hereby also acknowledge for generously supplying the project with material from their respective databases.

Ethical clearance for the research project was obtained through the Language Matters Ethics Committee of the NWU (ethics number: NWU-00632-19-A7). Additional ethics clearance for one of the master’s students was obtained from the Faculty of Humanities (UP), with reference number 16002360 (HUM017/0920).

The author is a director of the not-for-profit company Viridevert NPC (CIPC registration number: 2016/411799/08), who owns and manages the website vloek.co.za. This website was developed specifically for this project, and this conflict of interest has been approved by NWU.

I would like to acknowledge the inputs of Liesbeth Augustinus and Peter Dirix (Catholic University of Leuven, Belgium) in the initial conceptualization of this project, as well as Suléne Pilon (UP) in the ongoing re-conceptualization of the project. Thanks also to all the coworkers, collaborators, and students on the project; the list is too long to publish here, and it is therefore published on vloek.co.za/oor-ons.

None of the results and/or opinions in this paper can be ascribed to any of the people or organizations mentioned above.

References

Allan, K (ed.) 2019, *The Oxford Handbook of Taboo Words and Language*, Oxford University Press, Oxford.



- Beers Fägersten, K 2007, *A sociolinguistic analysis of swearword offensiveness*, Universität des Saarlands, Saarbrücken, view 22 August 2021, <https://www.researchgate.net/publication/265009714_A_sociolinguistic_analysis_of_swearword_offensiveness>.
- Beers Fägersten, K 2012, *Who's Swearing Now? The Social Aspects of Conversational Swearing*, Cambridge Scholars Publishing, Newcastle upon Tyne.
- Beers Fägersten, K & Stapleton, K (eds.) 2017, *Advances in Swearing Research: New Languages and New Contexts*, John Benjamins, Amsterdam.
- Bergen, BK 2016, *What the F: What Swearing Reveals About Our Language, Our Brains, and Ourselves*, Basic Books, New York.
- Calitz, FC 1979, Spot, skel en verwante verskynsels in Afrikaans [Mockery, swearing and related phenomena in Afrikaans], PhD thesis, Stellenbosch University, Stellenbosch.
- Coetzee, F 2018, 'Hy leer dit nie hier nie ('He doesn't learn it here'): talking about children's swearing in extended families in multilingual South Africa', *International Journal of Multilingualism*, vol. 15, no. 3, pp. 291-305.
- Dekker, L 1991, 'Vloek, skel en vulgariteit: Hantering van sosiolinguisties aanstootlike leksikale items', *Lexikos*, vol. 1, pp. 52-60.
- Eiselen, ER & Van Huyssteen, GB 2021, 'Using ordinal logistic regression to analyse self-reported usage of, and attitudes towards swearwords', *Proceedings of the International Conference of the Digital Humanities Association of Southern Africa 2021*, 29 November to 3 December, DHASA, South Africa.
- Feinauer, AE 1981, Die taalkundige gedrag van vloekwoorde in Afrikaans [The linguistic behaviour of swearwords in Afrikaans], MA dissertation, Stellenbosch University, Stellenbosch.
- Fillmore, CJ, Bernal, E & DeCesaris, J 2008, 'Border Conflicts: FrameNet Meets Construction Grammar', *Proceedings of the XIII EURALEX International Congress*, Universitat Pompeu Fabra, Barcelona.
- Global Language Monitor 2009, 'Death of Michael Jackson', view 22 August 2021, <<http://www.languagemonitor.com/news/death-of-michael-jackson/>>.
- Goel, R, Soni, S, Goyal, N, Paparrizos, J, Wallach, H, Diaz, F & Eisenstein, J 2016, 'The Social Dynamics of Language Change in Online Networks', International Conference on Social Informatics (SocInfo16), view 22 August 2021, <<https://arxiv.org/abs/1609.02075>>.
- Höder, S 2018, 'Grammar is community-specific. Background and basic concepts of Diasystematic Construction Grammar', in Boas, HC and Höder, S (eds.) *Constructions in Contact: Constructional perspectives on contact phenomena in Germanic languages*, Benjamins, Amsterdam, pp. 37-70.
- Hughes, G 2006, *An encyclopedia of swearing: the social history of oaths, profanity, foul language, and ethnic slurs in the English-speaking world*, M.E. Sharpe, Armonk.
- Jay, T 2000, *Why we curse: A neuro-psycho-social theory of speech*, John Benjamins, Amsterdam.
- Labov, W 2001, *Principles of Linguistic Change, Volume 2: Social Factors*, Language in Society, Wiley-Blackwell, London.
- Lavrakas, PJ 2008, *Encyclopedia of survey research methods* (Vols. 1-0), Sage Publications, Thousand Oaks, doi: 10.4135/9781412963947.
- Leroux, E 1976, *Magersfontein, O Magersfontein!*, Human & Rousseau, Cape Town.
- Leroux, E 1990, *Magersfontein: Die dokumente*, Human & Rousseau, Cape Town.
- Lyngfelt, B, Borin, L, Ohara, K & Torrent, TT (eds.) 2018, *Constructicography: Constructic development across languages*, John Benjamins, Amsterdam.
- McWhorter, JH 2021, *Nine nasty words: English in the gutter: then, now, and forever*, Kindle edn, Penguin, New York.
- Meho, LI 2007, 'The rise and rise of citation analysis', *Physics World*, vol. 20, no. 1, pp. 32-36.
- MyBroadband 2018, 'South Africa wasting money on funding unemployable humanities and social science university students', view 22 August 2021,



<<https://mybroadband.co.za/news/government/278877-south-africa-wasting-money-on-funding-unemployable-humanities-and-social-science-university-students.html>>.

O'Driscoll, J 2020, *Offensive Language: Taboo, offence and social control*, Bloomsbury, London.

Paradowski, MB & Jonak, Ł 2012, 'Diffusion of Linguistic Innovation as Social Coordination', *Psychology of Language and Communication*, vol. 16, no. 2, pp. 131-142.

Pizarro Pedraza, A 2018, *Linguistic Taboo Revisited: Novel Insights from Cognitive Perspectives*, Cognitive Linguistics Research [CLR], De Gruyter Mouton, Berlin.

Remler, D 2014, 'Are 90% of academic papers really never cited? Searching citations about academic citations reveals the good, the bad and the ugly', view 22 August 2021, <<https://dahliaremler.com/2014/04/09/are-90-of-academic-papers-really-never-cited-searching-citations-about-academic-citations-reveals-the-good-the-bad-and-the-ugly/>>.

Republic of South Africa, 2019, 'Films and Publications Act (65/1996): Classification guidelines for the classification of films, interactive computer games and certain publications', Film and Publication Board, Department of Communications, Government gazette, no. 42380, notice 539, 5 April 2019, pp. 21-59, view 22 August 2021, <http://www.gpwonline.co.za/Gazettes/Gazettes/42380_05-4_NationalGovernment.pdf>.

Sheidlower, J 2009, *The F-Word*, Kindle edn., Oxford University Press, Oxford.

Traugott, EC & Trousdale, G 2013, *Constructionalization and constructional changes*, Oxford University Press, Oxford.

Van der Walt, A 2019, Linguistiese eienskappe en konvensionalisering in Zefrikaans op die WatKykJy?-blog: 'n korpuslinguistiese ondersoek [Linguistic features and conventionalisation in Zefrikaans on the WatKykJy? blog: a corpus linguistic study], MA dissertation, North-West University, Vanderbijlpark.

Van Huyssteen, GB 1996, 'The sexist nature of sexual expressions in Afrikaans', *Literator*, vol. 17, no. 3, pp. 119-135.

Van Huyssteen, GB 1998, 'Die leksikografiese hantering van seksuele uitdrukkings in Afrikaans [The lexicographic treatment of sexual expressions in Afrikaans]', *South African Journal of Linguistics*, vol. 16, no. 2, pp. 63-71.

Van Huyssteen, GB & Eiselen, R 2021, 'Oor feekse en helleveë [On shrews and harridans]', *Tydskrif vir Geesteswetenskappe*.

Van Rooyen, K 2012, *A South African Censor's Tale*, Protea Boekhuis, Pretoria.

Van Sterkenburg, PGJ 2019, *Rot lekker zelf op: Over politiek incorrect en ander ongepast taalgebruik*, Scriptum, Schiedam.

VivA 2021, Virtual Institute for Afrikaans: Corpus Portal, view 19 August 2021, <<http://viva-afrikaans.org>>.

