

# Die rol van die Afrikaanse Wikipedia in die uitbou van Afrikaans

*The role of the Afrikaans Wikipedia in the growth of Afrikaans*

**LAURETTE PRETORIUS**

Skool vir Interdisiplinêre Navorsing en Nagraadse Studies

Universiteit van Suid-Afrika

E-pos: pretol@unisa.ac.za



Laurette Pretorius

**LAURETTE PRETORIUS** is professor in rekenaarwetenskap en lei die Unisa Strategiese Projek: Die Unisa Akademie vir Afrikatale en die Wetenskap, wat die intellektualisering van die Afrikatale op die terreine van die wetenskap, tegnologie, akademiese en onderwys deur middel van taaltegnologie ten doel het.

Sy het nagraadse kwalifikasies in rekenaarwetenskap, suiwer wiskunde en toegepaste wiskunde aan die Universiteite van Stellenbosch, Suid-Afrika, Pretoria en Potchefstroom verwerf. Haar navorsing is toegespits op die natuurliketaalverwerking (NTV) van die hulpbron-skaars Suid-Afrikaanse tale, insluitende Zoeloe, Xhosa, Tswana en Afrikaans. Haar navorsingsbelangstellings sluit ook die rol van natuurlike taal en veeltalige NTV in semantiese berekening en ontologie-ontwikkeling, beheerde natuurlike taal vir domein-spesifieke presiese masjienvertaling en taalgenerering in. Sy het reeds meer as 80 artikels in wetenskaplike tydskrifte en portuur-gekeurde konferensieverrigtinge gepubliseer, is sedert 2004 'n NRF-gegradeerde navorser en het tans 'n B-gradering.

Sy het uitgebreide navorsingbesoeke gebring aan die Xerox Navorsingsentrum Europa, Frankryk; Tilburg Universiteit, Nederland; die Departement Rekenaarwetenskap, Universiteit van Göteborg, Swede; die Insight Centre for Data Analytics, Nasionale Universiteit van Ierland, Galway; en die Universiteit van Helsinki, Finland.

**LAURETTE PRETORIUS** is a professor of computer science and leads the Unisa Strategic Project: The Unisa Academy of African Languages and Science, the aim of which is to intellectualise the African Languages in the scientific, technological, academic and educational domains by means of language technology.

She holds postgraduate degrees in computer science, pure mathematics and applied mathematics of the Universities of Stellenbosch, South Africa, Pretoria and Potchefstroom. Her research specialisation is the natural language processing (NLP) of the lesser-resourced Southern African languages, including Zulu, Xhosa, Tswana and Afrikaans. Her research interests also include the role of natural language and multilingual NLP in semantic computing and ontology development, controlled natural language for domain specific precise machine translation and language generation. She has published more than 80 articles in scientific journals and peer-reviewed conference proceedings, has been an NRF rated researcher since 2004 and currently holds a B-rating.

She has undertaken extended research visits to the Xerox Research Centre, Europe, France; Tilburg University, The Netherlands; the Department of Computer Science, University of Gothenburg, Sweden; the Insight Centre for Data Analytics, National University of Ireland, Galway; and the University of Helsinki, Finland.

**ABSTRACT*****The role of the Afrikaans Wikipedia in the growth of Afrikaans***

*In 2015, Afrikaans celebrated its 90th birthday as an official language of South Africa. In the course of these 90 years, Afrikaans progressed in the physical space from a so-called kitchen and domestic language to a mature, fully developed language, suitable and extensively used in all walks of life, including the church, school, university, science, sports, culture, law and economics. From being considered as “language of the oppressor”, Afrikaans grew into a model for the other South African languages; into a language of hope (Langner 2015). Influential writers refer to the electricity, the wonder, the triumph, but also the tragedy of Afrikaans (Steyn 2014; McLachlan 2010) due to the recently increasing pressure towards terminating the use of Afrikaans in secondary and higher education, and the work place. By 2015 it was clear that the efforts towards advancing Afrikaans were by no means a matter of the past.*

*Now, in its 91st year, Afrikaans is confronted with the #AfrikaansMustFall movement, which has manifested itself in both the physical and digital space (often also referred to as cyberspace). Here we consider digital space as the non-physical space within which we communicate by means of telecommunication and computer networks, more specifically the Internet, and the World Wide Web. In terms of the role that digital space and specifically the social media are playing in this movement, it resembles what has become known as the “Arab spring”, a movement that has radically changed the world. Similarly, the position of Afrikaans has been changed irrevocably by #AfrikaansMustFall. This raises the question of how to address this issue, also in the context of language planning.*

*The past 15 years have been characterised by an increasing migration of Afrikaans speakers into the digital space – a space that offers exciting new opportunities for Afrikaans.*

*In order to make a constructive contribution to the growth of Afrikaans, this article posits the following:*

- *that the focus until recently has been on the use and growth of Afrikaans in the physical space, but*
- *that the space in which (Afrikaans) speakers live has been radically changed by the advent of modern (personal) computers, mobile technology and ever-increasing and faster electronic networks, which has resulted in a shift towards life in the digital space;*
- *that every aspect of the Afrikaans speaker’s existence, also his/her communication and language, has been affected by this shift;*
- *that this trend is of key importance for future language diversity, and therefore also for Afrikaans;*
- *that the vitality of Afrikaans will be determined increasingly by its use in the digital space; and*
- *that the Afrikaans Wikipedia is a prerequisite for Afrikaans to ascend and flourish in the digital space.*

*It is shown that language resources and language technology are central to this endeavour. In particular, it is explicated how the Afrikaans Wikipedia plays a pivotal role in the digital vitality of Afrikaans and that a high-quality, growing Wikipedia is a prerequisite for Afrikaans and also any other language to ascend in the digital realm. We also discuss various possibilities for Afrikaans speakers to contribute to the Afrikaans Wikipedia.*

*The structure of the article is as follows: Following the introduction, Section 2 discusses the notion of language growth in the physical space. We consider various frameworks for studying*

*language vitality, focussing on the conceptual similarities between these frameworks. In Section 3 we explore the concepts of the digital space, the Internet and the World Wide Web, paying specific attention to the essential components of successful web search. In Section 4 we explore the extent to which South Africans live in the digital space. By means of an example, we illustrate the difference between the English and Afrikaans web search experience and the related significance of the size of the respective Wikipedias. We then briefly emphasise the ground-breaking work of András Kornai (2013), which shows that the existence of a large, high-quality, growing Wikipedia is a necessary condition for the digital ascent of a language. Section 5 is devoted to a brief summary of those language resources and language technologies that are necessary for the digital ascent and vitality of a language. The situation of Afrikaans is described in this regard. Section 6 discusses Wikipedia, the largest, multilingual, open and free online-encyclopaedia on the Web, with its more than 36 million articles, 292 languages and almost 500 million unique visitors per month. We briefly explain what Wikipedia is and how it plays a key role in the so-called Semantic Web, the intelligent machine-processable web. Section 7 focuses on the Afrikaans Wikipedia, provides short guidelines on how to contribute to a “digital language movement”, and concludes that the Afrikaans Wikipedia should play a vital role in the growth of Afrikaans in the digital space.*

**KEY CONCEPTS:** Afrikaans, Wikipedia, on-line encyclopaedia, language growth, digital language vitality, physical realm, digital realm, Internet, World Wide Web, Semantic Web, digital language resources, language technology, DBpedia, content translation, digital language movement

**TREFWOORDE:** Afrikaans, Wikipedia, aanlynensiklopedie, taaluitbou, digitale groeikrag, fisiese ruimte, digitale ruimte, Internet, Wêreldwye Web, Semantiese Web, digitale taalhulpbronne, taaltegnologie, DBpedia, inhoudsvertaling, digitale taalbeweging

## **OPSOMMING**

Afrikaans het in sy eerste 90 jaar as amptelike taal van Suid-Afrika van 'n sogenaamde kombuistaal tot 'n volledige onderwys-, wetenskaps- en kultuurtaal ontwikkel. Tog is Afrikaans in die fisiese ruimte waarin sy sprekers hulle bevind, toenemend onder druk en word sy amptelike gebruik weereens bedreig. Vanweë die tegnologiese voortuitgang op die gebied van telekommunikasie en rekenaarnetwerke, die Internet en die Wêreldwye Web het daar in die afgelope 15 jaar 'n digitale ruimte (kuberruimte) ontstaan waarbinne ook Afrikaanssprekendes toenemend leef. Hierdie artikel beskou die posisie van Afrikaans in hierdie digitale ruimte. Onlangse gesaghebbende navorsing deur András Kornai (2013) het aangetoon dat die bestaan van 'n groot, hoëgehalte-, groeiende Wikipedia 'n noodsaaklike voorwaarde is vir die digitale groeikrag van 'n taal. In hierdie artikel word die Afrikaanse Wikipedia van nader beskou. Ons toon aan dat dit nog nie omvangryk genoeg is om Afrikaans volledig in die digitale ruimte te vestig nie. Die rol van die Afrikaanse Wikipedia, beide as enorme oop en vrye aanlynensiklopedie en as spilpunt van die Semantiese Web en web-skaal taalhulpbronne en taaltegnologie, word uiteengesit. Praktiese riglyne vir hoe elke Afrikaanssprekende betrokke kan raak by die ontwikkeling van die Afrikaanse Wikipedia, word kortliks bespreek. Die artikel kom tot die slotsom dat 'n soort digitale taalbeweging met die ontwikkeling van die Afrikaanse Wikipedia as fokus nodig is om die posisie van Afrikaans in die digitale ruimte te verseker en vol te hou.

## 1. INLEIDING

In 2015 het Afrikaans sy 90ste verjaarsdag as amptelike taal van Suid-Afrika gevier. In hierdie 90 jaar het Afrikaans in ons fisiese wêreld (leefruimte) van sogenaamde kombuis- en huistaal ontwikkel tot volwaardige kansel-, skool-, universiteits-, navorsings-, sport-, kultuur-, regs- en ekonomiese taal. Afrikaans is al beskou as die taal van die onderdrukker, ’n model vir die ander Suid-Afrikaanse tale, en ook as ’n taal van hoop (Langner 2015). Afrikaans word, saam met moderne Hebreeus, beskou as een van die taalwonderwerke van die 20ste eeu (Prah 2006). Gesaghebbende skrywers het geskryf oor die elektrisiteit, die wonder, die triomf, die sieraad, maar ook die tragedie van Afrikaans (Steyn 2014; McLachlan 2010) vanweë toenemende druk op die gebruik van Afrikaans in skole, universiteite en die werkplek. Teen 2015 was daar dus steeds die suggestie van ’n uitboutsak wat nooit voltooi sal wees nie.

Nou, in sy 91ste jaar, is Afrikaans gekonfronteer met die #AfrikaansMustFall-beweging, ’n beweging wat hom in die fisiese sowel as die digitale leefruimte (kuberruimte) uitspeel. Ons beskou die digitale leefruimte as die (nie-fisiese) ruimte waarbinne ons deur middel van telekommunikasie- en rekenaarnetwerke, meer spesifiek die Internet, en die Wêreldwye Web, of kortweg die Web, kommunikeer. In terme van die rol wat die digitale ruimte en spesifiek die sosiale media daarin speel, herinner hierdie beweging sterk aan wat nou bekend staan as die “Arabiese lente”, ’n beweging wat die wêreld vir altyd ingrypend verander het. Afrikaans se posisie sal ook nooit weer wees wat dit voor #AfrikaansMustFall was nie. Die vraag wat ontstaan, is hoe om hierdie realiteit ten beste, ook binne die konteks van taalbeplanning, aan te spreek.

Die afgelope 15 jaar het ons toenemend in die digitale ruimte inbeweeg, en dit is hier waar daar ook opwindende nuwe geleenthede vir Afrikaans is. Ten einde ’n positiewe en konstruktiewe bydrae tot die posisie van Afrikaans te lewer, word die volgende standpunt in hierdie artikel gestel en begrond:

- dat die fokus tot baie onlangs op die gebruik en uitbou, die kwyn of groei, van Afrikaans in die *fisiese leefruimte* was – die huis, die werkplek, die kerk, die skool, die universiteit, die inkoopentrum, die bank, die hof, die parlement, die sportveld, die gedrukte media, die radio, ensovoorts; maar
- dat die ruimte waarin die Afrikaanssprekende leef, die afgelope 50 jaar radikaal verander het met die koms van die moderne (persoonlike) rekenaar, mobiele tegnologie, al groter en vinniger wordende elektroniese netwerke, ensovoorts, wat veroorsaak het dat ons ons toenemend ook in die sogenaamde *digitale leefruimte* bevind;
- dat elke faset van die Afrikaanssprekende se bestaan, en dus ook sy kommunikasie en taal, hierdeur geraak word;
- dat hierdie tendens van kardinale belang is vir taaldiversiteit in die toekoms, en dus spesifiek ook vir Afrikaans;
- dat die groeikrag van Afrikaans toenemend bepaal sal word deur sy gebruik in die digitale ruimte; en
- dat die Afrikaanse Wikipedia ’n noodsaaklikheid is vir Afrikaans om hierdie geleenthede te benut en in die digitale ruimte te floreer.

Ons toon dan aan dat taalhelpbronne en taaltegnologie sentraal staan in hierdie poging. In die besonder verduidelik ons dan hoedat die Afrikaanse Wikipedia ’n onontbeerlike spilpunt is vir die digitale groeikrag van Afrikaans – meer nog, dat sonder ’n groot, goeie en groeiende Wikipedia, Afrikaans en ook enige taal tot digitale sterfte gedoem is. Ons bespreek dan ten slotte hoe elke belangstellende spreker en gebruiker van Afrikaans ’n bydrae tot die Afrikaanse Wikipedia kan en behoort te lewer.

Die struktuur van die artikel is soos volg: Ná die inleiding, handel Afdeling 2 oor wat taaluitbou in die fisiese ruimte behels. Verskeie raamwerke vir die studie van die groeikrag van 'n taal word beskou, met spesifieke verwysing na die konseptuele ooreenkomste tussen hierdie raamwerke. In Afdeling 3 beskryf ons in meer besonderhede wat ons verstaan onder die digitale ruimte, die Internet, die Web en 'n suksesvolle websoektog. Afdeling 4 handel oor wat taaluitbou in die digitale ruimte behels. Aan die hand van 'n voorbeeld word die verskil tussen die Engelse en die Afrikaanse websoektog-ervarings geïllustreer en in verband gebring met die onderskeie groottes van die Engelse en Afrikaanse Wikipedias. Ons beklemtoon aan die hand van die baanbrekerswerk van András Kornai (2013) dat die bestaan van 'n groot, goeie en groeiende Wikipedia 'n voorvereiste is vir die voortbestaan van 'n taal in die digitale ruimte. Afdeling 5 bevat 'n kort opsomming van die taalhulpbronne en taaltegnologie wat belangrik is vir die groeikrag van 'n taal in die digitale ruimte, en beskou die posisie van Afrikaans in hierdie verband. Afdeling 6 handel oor Wikipedia, die grootste veeltalige, oop en vrye aanlynensiklopedie op die Web met sy meer as 36 miljoen artikels, 292 tale en bykans 500 miljoen unieke besoekers per maand. Ons verduidelik kortliks wat Wikipedia is en hoe Wikipedia 'n kernrol vervul in die sogenaamde Semantiese Web, die intelligente, masjien-leesbare web. In Afdeling 7 beskou ons die Afrikaanse Wikipedia van naderby, verskaf bondige riglyne oor hoe om betrokke te raak by 'n "digitale taalbeweging", en sluit af met 'n samevatting van die rol van die Afrikaanse Wikipedia in die uitbou van Afrikaans in die digitale ruimte.

## 2. TAALUITBOU IN DIE FISIESE RUIMTE

Ten einde sin te maak van die uitbou, of anders gestel, die groeikrag, van 'n taal in die digitale ruimte, is dit nuttig om eers te besin oor wat dit in die fisiese ruimte sou beteken. Die faktore waarvolgens die groeikrag van 'n taal bepaal word, is reeds omvattend nagevors (Obiero 2010). Vir die doeleindes van hierdie artikel bepaal ons ons by die mees algemene benaderings, te wete dié van UNESCO,<sup>1</sup> SIL International<sup>2</sup> en Kornai (2013), wat in wese heelwat ooreenstem:

UNESCO:

1. Tussengenerasietaaloordrag
2. Totale aantal sprekers
3. Verhouding sprekers in die totale bevolking
4. Verskuiwing in terreine van taalgebruik
5. Response tot nuwe terreine en media
6. Besikbaarheid van stof vir taalonderrig en geletterdheid
7. Regerings- en institusionele taalhouding en beleid, insluitend status en gebruik
8. Houding van die taalgemeenskap jeens hul eie taal
9. Hoeveelheid en kwaliteit van dokumentasie.

Aan die hand van UNESCO se nege faktore beskryf Dwyer (2011) hoedat hierdie faktore gebruik kan word om die groeikrag van 'n taal te bepaal. Terwyl dit relevant is om hierdie ontleding vir die groeikrag van Afrikaans in die fisiese ruimte te doen, val dit buite die bestek van hierdie artikel. SIL ondersoek die groeikrag van 'n taal aan die hand van die volgende:

<sup>1</sup> <http://www.unesco.org/new/en/culture/themes/endangered-languages/language-vitality/>

<sup>2</sup> <http://www.sil.org/language-assessment/language-vitality>

1. Funksies (terreine van gebruik) van 'n taal
2. Verwerwing (oordrag tussen generasies) van 'n taal
3. Motiewe en redes vir die gebruik van 'n taal
4. Regeringsbeleid aangaande taalgebruik
5. Spesifieke nisareas (bepaalde kontekste) waarin die taal gebruik word.

Die gesaghebbende Hongaarse rekenaartaalkundige, András Kornai, gebruik die volgende faktore (in volgorde van belangrikheid) in sy navorsing oor taalsterfte in die fisiese ruimte:

1. Grootte van die gebruikerspopulasie en die generasie-verspreiding: Die aantal en ouderdomme van die sprekers
2. Aansien: Die status en die nut wat die gebruik van 'n taal aan sy sprekers bied
3. Identiteitsfunksie: Die mate waartoe 'n spreker met 'n taal geassosieer wil wees en dit wil gebruik
4. Funksionele geskiktheid: Gebruiksomvang en gebruiksbevoegdheid van 'n taal
5. Funksionele verspreiding: Aktiewe gebruik oor geografiese, sosio-demografiese en ander grense heen.

Opsommend, hoewel daar verskillende konseptualiserings en formulerings is van die faktore wat 'n taal se groeikrag bepaal, blyk dit by nadere beskouing van die bostaande dat die groeikrag van 'n taal in die fisiese ruimte basies bepaal word deur sy aantal aktiewe sprekers, en veral die jonger generasies; die mate waartoe hulle die taal kan en wil gebruik; die mate waartoe die taal geskik is vir gebruik oor vele en veranderende terreine heen; en hoe wyd en diep die taal sigself reeds gevestig het. Omdat 'n taal en sy gebruikers so nou verbonde is aan mekaar, kom dit as geen verrassing nie dat soortgelyke faktore ook in die digitale ruimte geld.

### 3. DIE DIGITALE RUIMTE, DIE INTERNET EN DIE WÊRELDWYDE WEB

Om werklik die strategiese waarde van Wikipedia te kan beskryf, is dit nodig om 'n uiteensetting te gee van wat die digitale ruimte, die Internet en die Web is. Ons het reeds die *digitale ruimte* as die (nie-fisiese) ruimte waarbinne ons deur middel van telekommunikasie- en rekenaarnetwerke, meer spesifiek die Internet, en die Web kommunikeer, gedefinieer.

Kortweg gestel, is die Internet 'n reuse globale netwerk van rekenaarnetwerke waardeur reeds miljarde (teen 2020 meer as 50 miljard) rekenaars en digitale toestelle van oor die hele wêreld met mekaar geskakel is deur middel van telefoonlyne, kables, satellietverbindinge en ander netwerktegnologieë. Deur middel van die Internet is dit moontlik om enige rekenaar of toestel met enige ander een te verbind, op voorwaarde dat hulle aan die Internet gekoppel is. Die transmissie van inligting oor die Internet maak gebruik van 'n gestandaardiseerde suite protokolle bekend as TCP/IP (Transmission Control Protocol/Internet Protocol). Niemand besit die Internet nie. Verskeie organisasies regoor die wêreld, insluitend die telefoonmaatskappye van die onderskeie lande, werk egter saam aan die werking, instandhouding en ontwikkeling daarvan. Die Internet het tans meer as 3,3 miljard gebruikers, 45% van die wêreld se bevolking,<sup>3</sup> en is beskikbaar in 201 lande.<sup>4</sup> In Suid-Afrika met sy bevolking van 55 miljoen<sup>5</sup> gebruik 28,5 miljoen (52%) mense tans die Internet.<sup>6</sup> Die gebruike van die Internet word ruweg in drie groepe verdeel: vir

<sup>3</sup> <http://www.worldometers.info/world-population/>

<sup>4</sup> <http://www.internetlivestats.com/internet-users-by-country/>

<sup>5</sup> <http://www.worldometers.info/world-population/population-by-country/>

<sup>6</sup> <http://www.internetlivestats.com/internet-users-by-country/>



kommunikasie, vir lêeroordrag en vir die deel van inligting. Voorbeelde van kommunikasiedienste is elektroniese pos (e-pos), elektroniese poslyste, Internetgebaseerde telefonie, blitsboodskappe en aanlynklets, en voorbeelde van lêeroordragdienste is FTP (File Transfer Protocol) en Telnet (aanteken by 'n afgeleë rekenaar).

Die Web word ten beste beskryf as 'n model vir die deel van inligting wat bo-op die Internet gebou is en wat gebruik maak van HTTP (Hyper Text Transfer Protocol) om die funksies van skakel en blaai in die Web te bewerk en die kommunikasie tussen webbedieners en webgebruikers moontlik te maak. In meer praktiese terme is die Web 'n versameling teksbladsye, digitale foto's, musieklêers, video's, animasies, ensovoorts, wat in die vorm van sogenaamde webbladsye beskikbaar is oor die Internet. Elke webbladsy het 'n unieke identiteit, naamlik 'n sogenaamde URL (Uniform Resource Locator). 'n Webbladsy bevat tipies hiperteksskakels na ander webbladsye. Die Web gebruik programmatuur in die vorm van sogenaamde blaaiers soos Mozilla Firefox, Google Chrome, Internet Explorer, Opera en Safari om toegang tot hierdie webbladsye te kry. As 'n gebruiker reeds die URL van 'n bepaalde webbladsy ken, kan hy of sy dit in die soekvenstertjie van die blaaiër intik en so by die bladsy uitkom. Ten einde nuttige inligting vir die gebruiker op te spoor, maak die Web ook van soekenjins gebruik. Soekenjins soos Google, Yahoo en Bing is programmatuur wat die sleutelwoorde wat 'n gebruiker in die blaaiër se soekvenstertjie intik, gebruik om tersaaklike webbladsye te vind deur sistematies deur miljoene bladsye te soek wat in die betrokke soekenjin se databasis geïndekseer is. Die soekstrategie maak gebruik van patroonherkenning om die sleutelwoorde in tekste raak te sien. Soekenjins gebruik op hul beurt webspinnekoppe (programmatuur) om hierdie databasisse te bou. Spinnekoppe blaai outomaties en sistematies deur die Web en maak kopieë van al die webbladsye wat 'n soekenjin vir die doeleindes van vinniger soektogte in sy databasis wil indekseer. Kortom, vir die soek van inligting in die Web gebruik ons die blaaiers om by die soekenjins uit te kom, en die soekenjins om die gevraagde inligting te vind. Dit is dus duidelik dat dit die gesofistikeerdheid van die soekenjin is wat bepaal hoe goed of bruikbaar die inligting is wat die soektog oplewer. Ons kom terug na hierdie aspek wanneer ons die verskillende gedaantes van die Web bespreek.

Die mate waartoe ons toenemend in die digitale ruimte leef, staan in direkte verband tot die ontwikkeling van die Web. Web 1.0 (ongeveer 1991–1998) staan bekend as die statiese lees-Web, waarin die meeste Internet-gebruikers slegs die miljoene statiese webbladsye wat toe bestaan het, kon lees. Dit was 'n web van dokumente vir menslike verbruik en word beskou as die begin van die inligtingsontploffing. Groot hoeveelhede inligting was tegelyk altyd en onmiddellik beskikbaar. Daar was egter 'n duidelike onderskeid tussen die skeppers en die verbruikers van inligting, en geen kommunikasie was moontlik nie. Aanlyn-inkope dateer uit hierdie tydperk. Die keuse van taal was uitsluitlik die prerogatief van die eienaar van die inligting.

Die behoefte aan aktiewe interaksie en deelname van algemene gebruikers van die Web het tot Web 2.0 (ongeveer 1999–2008) gelei. Hierdie sogenaamde sosiale web, die dinamiese, lees-skryf-web waarin die skeidslyn tussen skepper en verbruiker vervaag het en almal in beginsel kon bydra, het 'n nuwe era ingelei. Die populariteit en die impak van die sosiale media, blogs, forums en aanlyn video-stroom het dit vir 'n ieder en 'n elk moontlik gemaak om met enkele klikke eie inhoud aanlyn te publiseer. Tipiese voorbeelde hiervan is Twitter, YouTube, Flickr, Instagram, Facebook, LinkedIn en natuurlik in 2001 Wikipedia (sien ook Afdeling 6). Web 2.0 was steeds net mens-leesbaar en taal was steeds die prerogatief van die skepper van die inligting. Hierdie web het dit egter ook vir Afrikaanssprekendes in beginsel moontlik gemaak om op groot skaal Afrikaanse inhoud te skep en te verbruik en so mee te leef in die digitale ruimte.

Web 3.0, ook bekend as die Semantiese Web (SW), is 'n verdere ingrypende uitbreiding van die Web soos dit hierbo beskryf is. Die SW is nie net mens-leesbaar nie, maar ook masjien-leesbaar,

d.w.s. leesbaar deur rekenaarprogramme, en staan ook bekend as die lees-skryf-uitvoer-web. Waar die Web 1.0 'n web van dokumente was, is die SW 'n web van data waar dit primêr gaan oor die semantiek (betekenis) van die data. Ten einde masjien-leesbaar te wees, word die SW gekenmerk deur 'n reeks tegniese standaarde, waarvan die belangrikste drie die volgende is:<sup>7</sup>

- (i). RDF (Resource Description Framework): Die datamodelleringstaal vir die SW. Alle inligting in die SW word in RDF gestoor en voorgestel.
- (ii). SPARQL (SPARQL Protocol and RDF Query Language): Die navraagtaal van die SW. Dit is spesifiek ontwerp om datanavrae oor 'n verskeidenheid (RDF) stelsels te doen.
- (iii). OWL (Web Ontology Language): Die skemataal of kennisvoorstellingstaal van die SW. Met OWL kan 'n mens konsepte noukeurig definieer, sodat hierdie konsepte maksimaal herbruikbaar is en op 'n verskeidenheid van maniere gekies en gekombineer kan word, soos wat verskillende toepassings en doelwitte dit vereis.

'n Uitgebreide verduideliking van hierdie standaarde val buite die bestek van hierdie artikel, maar die belangstellende leser word na Allemang en Hendler (2011) verwys. Ons keer terug na hierdie standaarde wanneer ons Wikipedia as spilpunt vir die SW bespreek.

Hierdie tegniese standaarde het dit ook moontlik en nodig gemaak vir soekenjins om “slimmer” te wees – om semanties te soek, in plaas van net patroonherkenning en sleutelwoorde te gebruik. In Afdeling 6 toon ons kortliks aan hoedat Wikipedia 'n sentrale rol hierin speel.

Die SW is van kardinale belang vir veeltaligheid in die Web. Die feit dat die SW masjien-leesbaar is, maak dit moontlik om, deur middel van die SW tegniese standaarde, veeltalige taaltegnologie te ontplooi wat oorkruisligheid moontlik maak, waar inligting wat in een taal beskikbaar is, ook vir gebruikers in ander tale toeganklik word – in effek 'n Veeltalige Semantiese Web (VSW), waarin Afrikaans ook teenwoordig is en groei.

Die Web ontwikkel steeds<sup>8</sup> en die moontlikhede is legio. Die Web van Dinge (Web of Things) bestaan reeds – 'n web waaraan nie net mense nie, maar ook slim toestelle deelneem deur inligting te gebruik, maar ook te skep. Die skatting is dat daar teen 2020 ses keer meer slim toestelle as mense aan die Web gekoppel sal wees. Deur die toepaslike taaltegnologie te ontwikkel, kan ons seker maak dat hierdie toestelle in Afrikaans met ons kommunikeer – nog 'n manier om Afrikaans in die digitale ruimte te vestig. Die Web word toenemend onsigbaar en alomteenwoordig soos die fisiese en die digitale ruimtes toenemend ineenvloei. Persoonlike toestelle soos selfone, yskaste en motors funksioneer buite die politieke sfeer van 'n beweging soos #AfrikaansMustFall, en gebruikers is dus vry om Afrikaanse taaltegnologie aan te wend in hulle daaglikse lewe.

#### 4. TAALUITBOU IN DIE DIGITALE RUIMTE

Voordat ons besin oor taaluitbou in die digitale ruimte, beskou ons kortliks Suid-Afrikaners se teenwoordigheid in die digitale ruimte. Suid-Afrika met sy bevolking van 55 miljoen (28 580 290 met Internettoegang) het naas Egipte (30 835 256 met Internettoegang) die tweede meeste Internetgebruikers in Afrika.<sup>9</sup> Ons buurstate het almal minder as 20%. Tabel 1 toon aan waarvoor Suid-Afrikaners hierdie toegang hoofsaaklik gebruik.<sup>10</sup> Uit hierdie tabel is dit duidelik dat ons reeds meeste van ons daaglikse aktiwiteite in die digitale ruimte uitvoer. Tabel 2 lys die mees

<sup>7</sup> <http://www.cambridgesemantics.com/semantic-university/introduction-semantic-web>

<sup>8</sup> <http://svicenter.com/technologyweb-3-0-semantic-web-future-internet-will-change-everything/>

<sup>9</sup> <http://www.internetlivestats.com/internet-users-by-country/>

<sup>10</sup> <http://mybroadband.co.za/news/internet/115575-what-south-africans-do-on-the-internet-4.html>



besoekte webblaaie in Suid-Afrika<sup>11</sup> en die wêreld,<sup>12</sup> onderskeidelik. Algemene soekenjins, sosiale media, vermaaklikheid, handel en loopbaannetwerke, en inderdaad Wikipedia, is almal teenwoordig en ondersteun die inligting in Tabel 1.

**TABEL 1:** Internetgebruik van Suid-Afrikanners in 2015

<b>Suid-Afrikanners se gebruik van die Internet</b>	<b>Persentasie Suid-Afrikanners wat dit benut</b>
E-pos	88.39%
Bankwese	65.60%
Navorsing / verkryging van inligting	62.32%
Nuus lees / aanlyntydskrifte lees	60.27%
Sosiale netwerke	54.56%
Inkope	34.42%
Werk soek	34.22%
Musiek aflaai	28.67%
Gidssoektogte	25.96%
Blitsboodskappe	23.08%
Klets	21.70%
Kyk TV en video's aanlyn	18.65%
Luister aanlyn radio	16.86%
Video's aflaai	14.18%
Blog	11.10%
Aandelehandel	7.05%
Sosiale afspraak	3.16%
Dobbel	1.64%
Geen van bogenoemde	1.17%

<sup>11</sup> <http://www.thespacestation.co.za/top-20-south-african-websites-for-september-2015/>

<sup>12</sup> <http://listabuzz.com/top-10-most-visited-websites-in-the-world/>

**TABEL 2:** Toptien webblaaie in Suid-Afrika en die wêreld in 2015

Rang	Suid-Afrika	Die wêreld
1	Google	Google
2	Google.co.za	Facebook
3	Facebook	YouTube
4	YouTube	Yahoo
5	Yahoo	Baidu.com
6	Gumtree.co.za	<b>Wikipedia</b>
7	Twitter	Tencent QQ
8	Amazon	LinkedIn.com
9	LinkedIn.com	Taobao.com
10	<b>Wikipedia</b>	Twitter

Kortom, dit is in die digitale ruimte waar ook die sprekers van Afrikaans toenemend deesdae hul banksake doen, aankope doen, sosiaal verkeer, musiek luister, films kyk, navorsing doen, koerant lees, korrespondeer en studeer. Tot nou was dit moontlik om al hierdie aktiwiteite in die fisiese ruimte in Afrikaans te verrig. Wat nou gemaak as die digitale ruimte stelselmatig besig is om die fisiese ruimte se funksies oor te neem? Ten einde Afrikaans te laat voortbestaan, sal die sprekers van Afrikaans die taal met hulle moet saamneem soos wat hulle toenemend in die digitale ruimte leef. Tans is dit nie die geval nie. Die webblaaie in Tabel 2 wat al die dienste verskaf en aktiwiteite in Tabel 1 ondersteun, gebruik hoofsaaklik Engels.

Ter illustrasie beskou ons die volgende tipiese voorbeeld:

Gestel iemand doen navorsing in die digitale ruimte en wil graag uitvind wie die vrou van John F Kennedy was. Die persoon gaan onmiddellik na Google, die nommer-een-webblad in Suid-Afrika en die wêreld, en tik die volgende vraag in die soekvenstertjie: “Who was the wife of John F Kennedy?” Die eerste wat aangebied word, is die inligting in Figuur 1: Die antwoord op die vraag, ’n foto van Jackie Onassis, verdere inligting oor haar en die bron van die inligting, naamlik die Engelse Wikipedia. Verder kry die navorser nie net die Wikipedia-bron nie, maar verskeie ander bronne (sien Figuur 2) waaruit die gegewe antwoord duidelik blyk. Die taak is suksesvol en vinnig afgehandel.

Gestel nou die navorser wil graag dieselfde navorsing in Afrikaans doen. Die persoon gaan weer na Google en tik die volgende vraag in die soekvenstertjie: “Wie was die vrou van John F Kennedy?” Wat eerste aangebied word, is ’n fragment uit ’n Huisgenoot-artikel wat deur Google geïndekseer is, soos in Figuur 3 getoon. Die antwoord is nie noodwendig daar nie, maar sleutelwoorde wat in die vraag voorkom, is beklemtoon. Verder bied die daaropvolgende trefslae van Google (sien Figuur 3) geen verdere lig op die saak nie. Die navorser moet dus verdere moeite doen deur nog ’n keer op die webadres <http://huisgenoot.com/nuus/charismatiese-president-se-dood-wereldwyd-herdenk/> te klik. Let op dat die antwoord toevallig in die fragment voorkom, maar nie beklemtoon is nie. Daarna moet die navorser die Huisgenoot-artikel lees om die antwoord op sy of haar vraag te vind. Dit is opvallend dat Google geen nuttige inligting in die Afrikaanse Wikipedia kon vind nie.

Dit is onwaarskynlik dat die ernstige Afrikaanse navorser, student of skolier tyd en geduld sal hê om in Afrikaans te soek, as die antwoord vinnig en akkuraat in Engels kom, maar die poging in Afrikaans lomp en stadig is.

John F. Kennedy / Spouse

---

## Jacqueline Kennedy Onassis

m. 1953–1963



---

Jacqueline Lee "Jackie" Kennedy Onassis was the wife of the 35th President of the United States, John F. Kennedy, and First Lady of the United States during his presidency from 1961 until his assassination in 1963. [Wikipedia](#)

---

More about Jacqueline Kennedy Onassis

*Figuur 1: Google se inligting, onttrek uit die Engelse Wikipedia*

**Jacqueline Kennedy Onassis - Wikipedia, the free ...**  
[https://en.wikipedia.org/wiki/Jacqueline\\_Kennedy\\_Onassis](https://en.wikipedia.org/wiki/Jacqueline_Kennedy_Onassis) ▾  
 Jump to Assassination and funeral of **John F. Kennedy** - **Jacqueline Lee "Jackie" Kennedy Onassis** (née Bouvier, pronounced /ˈdʒækliːn ˈliːˈbuːviɪr/; July 28, 1929 – May 19, 1994) was the wife of the 35th President of the United States, John F. Kennedy, and First Lady of the United States during his presidency from 1961 until his assassination in 1963.  
 Aristotle Onassis - Pink Chanel suit of Jacqueline - John Vernou Bouvier III - Lasata

**Life of Jacqueline B. Kennedy - John F. Kennedy ...**  
[www.jfklibrary.org](http://www.jfklibrary.org) > JFK ▾  
 In 1942, when Jackie was about to turn thirteen, her mother **married** a ... During this time, Jacqueline met **John F. Kennedy**, who was a congressman and soon to ...

**Jacqueline Kennedy Onassis - Publisher, U.S. First Lady ...**  
[www.biography.com/people/jacqueline-kennedy-onassis-9428644](http://www.biography.com/people/jacqueline-kennedy-onassis-9428644) ▾  
**Jacqueline Kennedy Onassis** was born **Jacqueline Lee Bouvier** on July 28, 1929, in Southampton, New York. She married John F. Kennedy in 1953.

*Figuur 2: Bevestiging van die Engelse Wikipedia se antwoord uit verskeie bronne*



**Figuur 3:** Google se inligting, onttrek uit 'n Huisgenoot-artikel

Dit blyk dus dat die stand van Afrikaans in die digitale ruimte tans nog heelwat te wense oorlaat. Die rol wat die Engelse Wikipedia in die Web speel, is steeds van uiterste belang omdat die Afrikaanse Wikipedia soos dit nou is, nie soektogte in Afrikaans ten volle ondersteun nie. Ons is nou gereed om taaluitbou en die groeikrag van 'n taal in die digitale ruimte van nader te beskou. Ons het reeds verwys na die navorsing van Kornai (2013 & 2015). Die faktore wat hy voorhou vir die groeikrag van 'n taal in die digitale ruimte, is 'n aanpassing en herformulering van die faktore vir die fisiese ruimte (soos in Afdeling 1 uiteengesit):

1. Grootte van die gebruikerspopulasie en die generasie-verspreiding: Aantal digitale boorlinge – aktief in digitale interaksie
2. Aansien: Geniet die taal digitale aansien en is dit digitaal nuttig?
3. Identiteitsfunksie: Kies gebruikers 'n bepaalde taal bo 'n ander in die digitale ruimte?
4. Funksionele geskiktheid: Die bestaan en wye beskikbaarheid van digitale taalhulpbronne en taaltegnologie, sodat die taal se gebruiksomvang en gebruikbevoegdheid in die digitale ruimte verseker is
5. Funksionele verspreiding: Die bestaan van 'n groot, hoëgehalte-, groeiende Wikipedia.

Die eerste drie faktore handel oor die gebruikers van Afrikaans in die digitale ruimte – oor hul aantal, ouderdom en houdings jeens die taal. Faktore vier en vyf handel oor hulpbronne wat in plek moet wees vir die genoemde gebruikers om Afrikaans suksesvol, toenemend en omvattend in die digitale ruimte te kan gebruik. Hoewel hierdie artikel nie in die eerste plek gaan oor faktor vier nie, hang dit tog nou saam met wat ons oor Wikipedia wil sê. Derhalwe gee ons in die volgende afdeling kortliks daaraan aandag.

In sy artikel “Digital Language Death” toon Kornai (2013) sistematies aan dat

Of the approximately 7,000 languages spoken today, some 2,500 are generally considered endangered. Here we argue that this consensus figure vastly underestimates the danger of digital language death, in that less than 5% of all languages can still ascend to the digital realm. We present evidence of a massive die-off caused by the digital divide.

...for the 95% of the world's languages where one or more of these drivers are missing, there is very little hope of crossing the digital divide

...To summarize a key result of this study ...: **No wikipedia, no ascent.**

Hy trek dus 'n duidelike lyn tussen die agteruitgang van 'n taal in die digitale ruimte en die afwesigheid van 'n florerende Wikipedia vir die taal. Dit is insiggewend om daarop te let dat sedert die publikasie van sy 2013-artikel, geen van sy bevindinge nog bevraagteken of weerlê is nie. Intendeel, verdere bevestiging en verfyning daarvan geniet tans aandag (Kornai 2015; Fenyvesi 2015; Soria 2015; Gibson 2015 & 2016).

Volgens Gibson (2015), word Kornai se assessering van digitale taalgroekrag, soos beskryf in sy “Digital Language Death”-artikel, as “the dominant model” beskou.

Soria (2015) skryf dat

Kornai’s “Digital language death” [Kornai 2013] represents the first attempt at devising reliable indicators of Digital Vitality ... by bringing the traditional methods of language vitality assessment to the digital realm. In doing so, he correctly identifies active digital uses of a language as a crucial factor in determining its Digital Vitality, and therefore suggests to complement the indicators of digital presence of a language (i.e. number of web pages in a given language) with other proxies for digital language use, such as the existence of an active Wikipedia community in the language.

Fenyvesi (2015) skryf

In a paper meticulously supported by ample empirical data and mathematical calculations, Kornai [2013] has argued that digital language death will be the fate of a great number of the languages existing today – primarily those that exist as minority languages only – unless their speakers (and the professionals supporting them) succeed in meeting some [or] all important criteria like having a community of digitally literate users and a Wikipedia in the language.

## 5. DIGITALE TAALHULPBRONNE EN TAALTEGNOLOGIE

Funksionele geskiktheid, faktor vier van Kornai (2013) vir die groekrag van 'n taal in die digitale ruimte, het te make met die bestaan en wye beskikbaarheid van digitale taalhulpbronne en taaltegnologie sodat die taal se gebruiksomvang en gebruiksbevoegdheid in die digitale ruimte verseker is. Onder taalhulpbronne verstaan ons<sup>13</sup> 'n versameling spraak- of teksdata en -beskrywings in masjien-leesbare formaat wat gebruik word vir die

- bou, verbetering of evaluering van algoritmes en stelsels vir natuurliketaalverwerking of spraak;
- as kernhulpbronne vir programmatuurlokaliseringsstelsels, spraakstelsels en die taaltegnologie-industrie;
- taalstudies, elektroniese publikasie, internasionale transaksies, domeinkenners en ander gebruikers.

Voorbeelde van sodanige taalhulpbronne is geskrewe en gesproke korpora (datastelle), masjien-leesbare leksika, terminologiedatabasisse, spraakversamelings, ensovoorts. Ten einde hierdie hulpbronne ten beste aan te wend, is basiese programmatuur noodsaaklik vir die verkryging, voorbereiding, versameling, bestuur, aanpassing en gebruik van hierdie hulpbronne. Een van die belangrikste verskille tussen wêreldtale met groot kommersiële reikwydte en kleiner tale is die wye beskikbaarheid, die hoeveelheid en die kwaliteit van taalhulpbronne. Sonder hierdie bronne is geen betekenisvolle taaltegnologie moontlik nie.

<sup>13</sup> <http://www.elra.info/en/about/what-language-resource/>

Taaltegnologie verwys na rekenaarmatige metodes, rekenaarprogramme en elektroniese artefakte wat toegespits is op die analise, produksie en die wysiging van taalktekste en spraakuitinge, en wat op die een of ander wyse gegrond is in die kennis van menslike taal. Van die belangrikste toepassings van taaltegnologie is sekerlik taaltoetsers, websoektoegte, masjienvertaling en spraakinteraksie. Tegnologieë wat tipies hiervoor nodig is, sluit in speltoetsers, outeurerondersteuning, rekenaargesteuende taalaanleer, inligtingherwinning, inligtingonttrekking, teksopsomming, vraag-antwoordsisteme, spraakherkenning en spraaksintese, wat op hulle beurt weer 'n verskeidenheid basiese komponente benodig (Ananiadou, McNaught & Thompson 2012).

'n Uitgebreide bespreking van taalhulpbronne en taaltegnologie vir Afrikaans val buite die bestek van hierdie artikel, maar die belangstellende leser kan meer inligting hieroor vind in byvoorbeeld, (Beesley, Bosch & Pretorius 2013; Grover et al. 2010).

Kornai beskryf 'n drievlakhiërargie as minimale vereistes vir taaltegnologie, naamlik (i) die beskikbaarheid van toevoer- en afvoermetodes vir die alfabet en ander Afrikaansspesifieke karakters; (ii) die beskikbaarheid van 'n aantal woordvlakhulpmiddels soos elektroniese woordeboeke, woordafbrekers en speltoetsers, weliswaar en ongelukkig nie almal oop en vryelik beskikbaar nie; (iii) Frase- en sinsvlakhulpmiddels soos woordsoortetiketteerders, benoemde-entiteitsherkenners, vlaksinsontleders, spraakherkenning en masjienvertaling. In terme hiervan is Afrikaans goed geposisioneer: In Maart 2016<sup>14</sup> is Afrikaans, op grond van die nuutste beskikbare data, as digitaal “Lewenskragtig” bevind, tweede op die vierpuntskaal van “Florerend” (Thriving), “Lewenskragtig” (Vital), “Erfenis” (Heritage) en “Stil” (Still) (Kornai 2013). Tans is Afrikaans een van bykans 250 tale wat die potensiaal het om in die digitale ruimte te kan oorleef.

## 6. WIKIPEDIA

### 6.1 Wikipedia as ensiklopedie

Wikipedia is in 2001 deur die Wikimedia Stigting<sup>15</sup> gestig met die volgende visie:<sup>16</sup> “Imagine a world in which every single human being can freely share in the sum of all knowledge. That’s our commitment.” Vandag, ná 15 jaar, is Wikipedia steeds die beste voorbeeld van 'n suksesvolle, dinamiese, veeltalige, sosiale webtoepassing. Met sy meer as 39 miljoen artikels oor 'n wye verskeidenheid van onderwerpe is Wikipedia die grootste, vrye-inhoud-,<sup>17</sup> algemene kennis-, webgebaseerde ensiklopedie ooit. Wikipedia verteenwoordig 292 tale, en het bykans 500 miljoen unieke besoekers per maand. Die Afrikaanse Wikipedia is op 1 November 2001 gestig as die elfde Wikipedia en is tans die 86ste grootste.

Die sukses van Wikipedia is sekerlik eerstens toe te skryf aan die feit dat dit as 'n gemeenskapsprojek gekonseptualiseer is wat volledig oop, vry en self-regulerend is. Die volledige geskiedenis van elke artikel in Wikipedia is vryelik beskikbaar, Wikipedia stel gereeld statistiek beskikbaar oor die deelname daaraan, vordering daarmee, ensovoorts. Deelnemers kan as anonieme gebruikers bydra, maar kan ook gebruikersname kies. Daar is ook stelsels in plek om vandalisme aan te spreek. As ensiklopedie word die gebruik van wetenskaplike verwysings in artikels aangemoedig. Nog 'n belangrike aspek van Wikipedia is dat sy ensiklopediese aard meebring dat die taalgebruik oor die algemeen van 'n hoë gehalte en register is – anders as ander sosiale media.

<sup>14</sup> Persoonlike kommunikasie van Andras Kornai op 22 Maart 2016

<sup>15</sup> <https://wikimediafoundation.org/wiki/Home>

<sup>16</sup> <https://wikimediafoundation.org/wiki/Vision>

<sup>17</sup> <https://en.wikipedia.org/wiki/Wikipedia:Copyrights>



Hoe betroubaar is Wikipedia as ensiklopedie? Omdat enigiemand inligting of artikels kan bydra, redigeer en verander, is die argument dikwels dat mense onbetroubare inligting daarop beskikbaar sou kon maak. Dit is natuurlik waar, maar daar is ook die teenargument dat 'n volgende persoon dit ook weer kan regstel. Die realiteit is dat dit met tyd, en met vele toegewyde en kundige medewerkers, beter word. Goeie Wikipedia-artikels het gewoonlik ook baie verwysings wat nagevolg en gebruik kan word om die inligting te verifieer. Die Engelse Wikipedia bevat 'n uitgebreide artikel, getiteld “Reliability of Wikipedia”,<sup>18</sup> met 237 verwysings waarin die kwaliteit toenemend bevestig word. Hierdie tema het dus in sigself 'n navorsingsveld geword.

Hoe omvattend is Wikipedia as ensiklopedie? Teen Maart 2016 het die Engelse Wikipedia 5, 105 616 artikels van goeie gehalte gehad, teenoor die 39 239 artikels van die Afrikaanse Wikipedia, waarvan talle kort en sommige van swakker gehalte was. Daarteenoor het die Hebreeuse Wikipedia 186 663 artikels en die Katalaanse 500 815 gehad<sup>19</sup> – dit terwyl Afrikaans meer moedertaalsprekers (6,9 miljoen) as beide Hebreeus (5 miljoen) en Katalaans (4 miljoen) het. Dat waardevolle werk oor die afgelope vyftien jaar aan die Afrikaanse Wikipedia gedoen is, ly geen twyfel nie. Dat daar vele artikels en 'n groeiende aantal van hoogstaande gehalte is, is sekerlik waar, maar as Afrikaans in groeikrag in die digitale ruimte moet aanhou toeneem, lê hef aan vir die Afrikaanse Wikipedia nog voor.

## 6.2 Wikipedia as spilpunt vir die Semantiese Web en taaltegnologie

Wikipedia het meer as 'n ensiklopedie geword. Dit is geen wonder dat die grootste, veeltalige, webgebaseerde, vrye-inhoud-kennisbron die verbeelding van twee verskillende, maar tog aanverwante vakgemeenskappe aangegryp het nie: (i) die Semantiese Web-gemeenskap wat hom besig hou met kennisvoorstelling, intelligente inligtingonttrekking, semantiese soek en soekenjins; en (ii) die taaltegnologie-gemeenskap wat voortdurend op soek is na taaldata (die Web en Wikipedia as korpus) vir navorsing in natuurliketaalverwerking en, meer spesifiek dan, parallelle korpora (taaldata in meerdere tale) vir masjienvertaling. Hierdie twee vakgemeenskappe het mekaar in die konteks van die Veeltalige Semantiese Web (VSW) gevind. Dit is juis die VSW wat veeltaligheid in die digitale kennisontploffing moontlik maak, wat ook as digitale platform vir kulturele en taaldiversiteit dien, en as een van die belangrikste moderne tegnologiese ontwikkelings vir die demokratisering van kennis – 'n sogenaamde virtuele kennismeent – beskou word (Pretorius2014).

Maar hoe presies het Wikipedia die spilpunt vir veeltalige digitale taalhulpbronne en tegnologieë, asook webskaal-kennisstelsels en semantiese soek geword?

Net soos Wikipedia, is DBpedia<sup>20</sup> 'n gemeenskapsprojek wat ten doel het om eksakte gestruktureerde inligting uit Wikipedia te onttrek en op die Web beskikbaar te stel. Dit maak gebruik van die drie tegniese tale van die SW, wat in Afdeling 3 bespreek is (RDF, SPQL en OWL), om die inligting te struktureer en beskikbaar te stel vir die gevorderde algoritmes en rekenaarsstelsels wat die SW en die moderne soekenjins onderlê. DBpedia maak dit moontlik om komplekse navrae aan Wikipedia te rig, en ook ander datastelle met die Wikipedia-data te skakel, om sodoende verdere waarde toe te voeg aan die Wikipedia-data. As enorme databasis maak DBpedia dit moontlik om die steeds groeiende versameling van kennis in Wikipedia beskikbaar te maak onder 'n vry en oop lisensie, soortgelyk aan dié van Wikipedia. Die volledige DBpedia-databasis bevat tans 38 miljoen etikette en opsommings in 125 verskillende tale, meer as 25 miljoen

<sup>18</sup> [https://en.wikipedia.org/wiki/Reliability\\_of\\_Wikipedia](https://en.wikipedia.org/wiki/Reliability_of_Wikipedia)

<sup>19</sup> [https://en.wikipedia.org/wiki/List\\_of\\_Wikipedias](https://en.wikipedia.org/wiki/List_of_Wikipedias)

<sup>20</sup> <http://wiki.dbpedia.org/>





## 7. 'N DIGITALE TAALBEWEGING?

Kan ons praat van 'n digitale taalbeweging? “'n Taalbeweging is ... nie alleen 'n stryd om bepaalde 'hoë funksies' vir die taal te verower nie, maar om die taal ook geskik te maak vir dié funksies” (Steyn 1980:211). 'n Volledige uiteensetting van die Afrikaanse taalbewegings sedert 1875 is buite die bestek van hierdie artikel. Ons beklemtoon egter graag enkele aspekte van hierdie taalbewegings ten einde aan te toon dat ons in hierdie artikel te make het met 'n digitale taalbeweging met die Afrikaanse Wikipedia as kern. Die Eerste Taalbeweging is onder andere gekenmerk deur die skep van oorspronklike Afrikaanse tekste oor 'n wye verskeidenheid terreine (Steyn 1980:139) en die totstandkoming van *Die Afrikaanse Patriot*, die eerste Afrikaanstalige koerant. Steyn (1980:211-218) gee verder 'n kompakte opsomming van die onskatbare betekenis en omvang van Afrikaanse skryfwerk wat gedurende die Tweede Taalbeweging (sedert 1905) plaasgevind het. Wanneer Webb (2010) egter skryf oor 'n Derde Taalbeweging sedert 2003, verwys hy hoofsaaklik na georganiseerde pogings om diegene wat Afrikaans praat, te verenig tot “a single inclusive community, which should then become a cohesive entity and possess power”. Hierdie taalbewegings, al drie in die fisiese ruimte, het dus gefokus op die grootskaalse skep van Afrikaanse inhoud deur en vir 'n bepaalde taalgemeenskap, die Afrikaanssprekendes. In hierdie opsig toon dit 'n duidelike ooreenkoms met die skep van 'n groot en groeiende Afrikaanse Wikipedia deur en vir die totale Afrikaanssprekende gemeenskap, inderdaad 'n digitale taalbeweging.

Wat is dus nodig vir Afrikaans om deur middel van sy Wikipedia in die digitale ruimte te gedy? Die kort antwoord is dat die Afrikaanse Wikipedia aansienlik en voortdurend uitgebrei moet word. Wikipedia is 'n gemeenskapsprojek, en as sodanig is dit dus die verantwoordelikheid van al die sprekers van 'n taal om daaraan mee te doen.

Die alternatief om professionele vertalers te gebruik, kan slegs 'n korttermynmaatreël wees en is helaas nie volhoubaar nie. Dit blyk duidelik uit die volgende kosteberekening: Veronderstel argumentsonthalt dat die doelwit is om teen 2025, wanneer Afrikaans as amptelike taal 100 jaar oud is, 100 000 artikels in die Afrikaanse Wikipedia te hê. Dit sou beteken dat nog ongeveer 60 000 artikels toegevoeg moet word. As ons reken dat 'n artikel gemiddeld 1 500 woorde lank moet wees (die artikel oor FW de Klerk is 1 822 woorde lank, en die een oor Nelson Mandela 1 275 woorde) en dat die Suid-Afrikaanse Vertalersinstituut (SATI) tans 'n tarief van 65 sent per woord aanbeveel, beloop die koste R975 per artikel, en sal dit R58,5 miljoen kos om 60 000 artikels te vertaal. In die huidige ekonomiese klimaat is dit nouliks haalbaar.

Die Wikimedia Stigting wat die programmatuur vir Wikipedia onderhou, het 'n geïntegreerde vertaalomgewing<sup>26</sup> (Content Translation (CX))<sup>27</sup>, spesifiek vir die vertaling van Wikipedia-artikels, ontwikkel (Laxström, Giner & Thottingal 2015).

Dit hou verskeie voordele in:

- Dit is maklik om te gebruik;
- Dit is met Wikipedia geïntegreer, met behoud van uitleg, beeldmateriaal, wiskundige formules, skakels, ensovoorts – alles sigbaar tydens vertaling;
- Dit het 'n koppelvlak soortgelyk aan wat Wikipedia self gebruik vir redigering. Dus sal vertalers dieselfde koppelvlak hier kan gebruik as wat hulle tydens redigeerwerk op Wikipedia benodig;
- Dit voorkom duplisering van artikels, deurdat dit reeds die skakeling van bestaande artikels met Engelse brontekste deur middel van Wikidata ken;

<sup>26</sup> [https://www.mediawiki.org/wiki/Content\\_translation](https://www.mediawiki.org/wiki/Content_translation)

<sup>27</sup> <http://blog.wikimedia.org/2015/04/08/the-new-content-translation-tool/>



- Dit verskaf konteksinsligting vanuit Wikipedia self, wat onder andere help met die korrekte hantering van skakels;
- Dit maak dit makliker om aan die konvensies en vereistes van Wikipedia te voldoen – só word die risiko van weerstand van die Wikipedia-gemeenskap verlaag, en die kanse op positiewe samewerking met bestaande medewerkers verhoog;
- Dit behou vertalings outomaties vir ander en latere gebruik. Dit is 'n langtermynbydrae tot Wikipedia en die wyer gemeenskap vir navorsing;
- Dit word wyd gebruik in vele tale in Wikipedia, en die programmatuur word deur die Wikimedia Stigting onderhou en verder ontwikkel;
- Dit hanteer die aanpassing van skakels outomaties. Kategorieë word sover moontlik outomaties aangepas vir die Afrikaanse stelsel van kategorisering, en verminder só handmatige werk wat agterna gedoen moet word;
- Dit bied reeds ondersteuning vir sjablone, verwysings en tabelle wat algemeen voorkom in langer, meer komplekse artikels;

Die inhoudsvertalingstelsel het tans enkele nadele, maar die stelsel word deurlopend verbeter en die verwagting is dat die onderstaande nadele in die afsienbare toekoms uit die weg geruim sal word:

- Dit maak nie voorsiening vir die vertaler om sy eie vertaalgeheue en terminologie te gebruik nie;
- Dit vereis 'n internetverbinding;
- Slegs webblaaiers se ingeboude speltoetsers is beskikbaar.

Die Afrikaanssprekende gemeenskap sal dus gratis en verniet skouer aan die wiel moet sit. 'n Taalbeweging is inderdaad nodig waarin kundige Afrikaanssprekendes op groot skaal self artikels vertaal deur van die CX Inhoudvertaling (soos hierbo beskryf) gebruik te maak, self redigeer, skryf en terminologie ontwikkel. Die Afrikaanssprekende gemeenskap is ideaal toegerus om te verseker dat Afrikaans ook sy status as taalwonderwerk in die digitale ruimte handhaaf. In hierdie verband het die Suid-Afrikaanse Akademie vir Wetenskap en Kuns onlangs 'n projek geloods om vrywillige kundiges touwys te maak oor hoe om tot die Afrikaanse Wikipedia by te dra.

## 8. SLOT

Onlangse navorsing het getoon dat Afrikaans een van die ongeveer 250 tale wêreldwyd is wat nog die potensiaal het om digitaal te bestaan en te gedy. Ook in die digitale ruimte is Afrikaans dus nog 'n taal van hoop. 'n Noodsaaklike voorwaarde hiervoor is egter dat Afrikaans 'n groot, groeiende en hoëgehalte-Wikipedia het – 'n gegewe wat ons verduidelik deur Wikipedia se rol as enorme oop en vrye aanlyn ensiklopedie te bespreek, en ook te toon hoe Wikipedia, as spilpunt van die Semantiese Web en web-skaal taalhulpbronne en taaltegnologie, die Veeltalige Semantiese Web moontlik maak. Verder is ook aangetoon dat die Afrikaanse Wikipedia egter nog nie omvangryk genoeg is om Afrikaans volledig in die digitale ruimte te vestig nie. Praktiese riglyne vir hoe Afrikaanssprekendes betrokke kan raak by die ontwikkeling van die Afrikaanse Wikipedia is vervolgens kortliks bespreek. Ons kom tot die slotsom dat 'n soort digitale taalbeweging waarin die Afrikaanse Wikipedia eweneens sentraal staan, nodig is om die posisie van Afrikaans in die digitale ruimte te verseker en vol te hou.

## BIBLIOGRAFIE

- Allemang, D. & Hendler, J. 2011. *Semantic web for the working ontologist*. 2de uitgawe. Amsterdam: Elsevier.
- Ananiadou A., McNaught J. & Thompson, P. 2012. *The English language in the digital age*. In: G. Rehm & H. Uszkoreit, eds. *White paper series*. Berlyn: Springer.
- Beesley, K.R., Bosch, S. & Pretorius, L. 2013. The impact of language technologies on South Africa's lesser-studied official languages. *Tromsø International Conference on Language Diversity*, 6-8 November 2013, Invited Thematic Workshop on 'The Impact of Language Technology on Society', 8 November 2013, Noorweë.
- Chiaros, C., Hellmann, S. & Nordhoff, S. 2012. Linking linguistic resources: Examples from the Open Linguistics Working Group, In: Christian Chiaros, Sebastian Nordhoff and Sebastian Hellmann (eds). *Linked Data in Linguistics. Representing Language Data and Metadata*. Heidelberg: Springer, pp. 201-216.
- Dwyer, A.M. 2011. Tools and techniques for endangered-language assessment and revitalization. In: *Trace Foundation Lecture Series Proceedings*. New York: Trace Foundation. Aanlyn: [http://www.trace.org/events/events\\_lecture\\_proceedings.html](http://www.trace.org/events/events_lecture_proceedings.html)
- Fenyvesi, A. 2015. Multilingualism and minority language use in the digital sphere: the digital use of language as a new domain of language use. In: *Proceedings of the 3rd International Conference on Linguistic and Cultural Diversity in Cyberspace*. Yakutsk, Russiese Federasie, Moskou, 30 Junie – 3 Julie 2014.
- Gibson, M. 2015. A framework for measuring the presence of minority languages in cyberspace. In: *Proceedings of the 3rd International Conference on Linguistic and Cultural Diversity in Cyberspace*. Yakutsk, Russiese Federasie, Moskou, 30 Junie – 3 Julie 2014.
- Gibson, M. 2016. Assessing digital vitality: analytical and activist approaches. In: *Proceedings of CCURL 2016, 2nd workshop on collaboration and computing for under-resourced languages, 'Towards an alliance for digital language diversity'*, 10th International Conference on Language Resources and Evaluation, LREC 2016, Portorož, Slovenia.
- Grover, A.S., Van Huyssteen, G.B. & Pretorius, M.W. 2011. The South African Human Language Technology Audit. *Language Resources and Evaluation*, 45(3), pp. 271-288.
- Kornai, A. 2013. Digital Language death. *PLoS ONE*, 8(10): e77056. doi:10.1371/journal.pone.0077056.
- Kornai, A. 2015. A New Method of Language Vitality Assessment. In: *Proceedings of the 3rd International Conference on Linguistic and Cultural Diversity in Cyberspace*. Yakutsk, Russiese Federasie, Moskou, 30 Junie – 3 Julie 2014.
- Langner, D. 2015. Afrikaans 90: Afrikaans is 'n taal van hoop. Aanlyn: <http://maroelamedia.co.za/afrikaans/afrikaans-90-afrikaans-is-n-taal-van-hoop/>
- Laxström, N., Giner, P. & Thottingal, S. 2015. Content translation: computer-assisted translation tool for Wikipedia articles. Aanlyn: <http://arxiv.org/abs/1506.01914>
- Margaretha, E. & Lungen, H. 2014. Building linguistic corpora from Wikipedia articles and discussions. *Journal for Language Technology and Computational Linguistics*, 29(2):59-82.
- McLachlan, T. 2010. Die groei van Afrikaans sedert 1910 tot 'n volwaardige taal. Aanlyn: <http://www.standpunte.co.za/standpunte/>
- Obiero, O.J. 2010. From assessing language endangerment or vitality to creating and evaluating language revitalization programmes. *Nordic Journal of African Studies*, 19(4):201-226.
- Prah, K.K. 2006. *Challenges to the Promotion of Indigenous Languages in South Africa*. Kaapstad: The Center for Advanced Studies of African Society.
- Pretorius, L. 2014. The Multilingual Semantic Web as Virtual Knowledge Commons: The Case of the Under-Resourced South African Languages. In: P. Buitelaar & P. Cimiano, eds. *Towards the Multilingual Semantic Web: Principles, Methods and Applications*. Heidelberg: Springer.
- Soria, C. 2015. Towards a notion of "Digital Language Diversity". In: *Proceedings of the 3rd International Conference on Linguistic and Cultural Diversity in Cyberspace*. Yakutsk, Russiese Federasie, Moskou, 30 Junie – 3 Julie 2014.
- Steyn, J.C. 1980. *Tuiste in eie taal: die behoud en bestaan van Afrikaans*. Kaapstad: Tafelberg.
- Steyn, J.C. 2014. *'Ons gaan 'n taal maak' – Afrikaans sedert die Patriot-jare*. Centurion: Kraal Uitgewers.
- Webb, V. 2010. Constructing an inclusive speech community from two mutually excluding ones: the third Afrikaans language movement. *Tydskrif vir Letterkunde*, 47(1):106-120.