

Chapter 2

A visit to SADiLaR

Una visita a SADiLaR
Menno Van Zaanen

SCAN THIS QR CODE TO WATCH A VIDEO ABOUT THIS CHAPTER
ESCANEE ESTE CODIGO QR PARA VER UN VÍDEO SOBRE ESTE CAPÍTULO

<https://www.youtube.com/watch?v=7zaU82f1piY>



A visit to SADiLaR

Menno Van Zaanen

South African Centre for Digital Language Resources (SADiLaR)

menno.vanzaanen@nwu.ac.za

Abstract

The South African Centre for Digital Language Resources (SADiLaR), funded by the South African Department of Science and Innovation, is a South African research infrastructure. The research infrastructure consists of a hub and nodes at different organizations (universities, an R&D organization, and an inter-institutional centre) throughout the country. SADiLaR runs two programmes: a digitization programme which deals with the creation and distribution of digital language resources (including written, spoken, and multimedia language collections as well as tools) for all of the eleven official South African languages, and a digital humanities programme which facilitates research capacity building in the field of digital humanities.

Even though many researchers in the field of humanities and social sciences are interested in digital humanities, many do not really know where to start. To try to resolve this, SADiLaR organizes a range of activities, which recently have been structured within a new programme called Escalator. One of these activities deals with showcasing research that illustrates the possibilities of digital humanities in an accessible manner. Here, we illustrate investigations into systems that may be used as showcases and at the same time can be used to visualize information that is contained in (specifically speech-based) digital language collections that also contain geographic information. SADiLaR considers the use of several tools, two of which are described in this chapter: Spraakatlas, and Glossopticon Virtual Reality (VR). Additionally, a third tool, Byderhand, which provides access to multimodal documents at specific locations in the physical world, is discussed as well. Currently, only exploratory investigations into these tools have taken place.

Keywords: Digital language resources, visualization, Spraakatlas, Glossopticon VR, Byderhand.

Introduction

The South African Centre for Digital Language Resources (SADiLaR) is a government funded project within the South African Research Infrastructure Roadmap (SARIR) programme (SARIR, 2016). The centre consists of a hub, located at North-West University in Potchefstroom, South Africa, and six nodes: NWU's Centre for Text Technology (CTextT) dealing with computational linguistic tools, ICELDA (an inter-institutional centre) focusing on language testing, University of Pretoria working on digitization, University of South Africa researching terminology and Wordnets, Council for Scientific and Industrial Research (CSIR) acting as the speech node, and recently Stellenbosch University joined as the child language development node. SADiLaR runs two programmes, a digitization programme and a digital humanities programme. These programmes are supported through funding (provided by SADiLaR for projects at the different nodes, but SADiLaR also funds projects through open calls) as well as practical support (e.g., organization of training events, conferences, etc.) by researchers at the hub. These programmes aim to provide resources that are openly accessible, although they are mostly relevant for students and researchers in the broad fields of humanities and digital humanities.

Digitization Programme

SADiLaR's digitization programme deals with the systematic creation of digital language resources, including the creation of text, speech, and multimodal collections for all of the eleven official languages in South Africa. This is performed, for instance, by SADiLaR's digitization node at the University of Pretoria, although additional projects that are funded by SADiLaR also result in language data collections. Additionally, this programme develops tools that allow for computational linguistic analysis of these languages. For instance, the Centre for Text Technology (CTextT), one of SADiLaR's nodes provides many of the available tools.

The different resources are made available through a repository¹. Identifying resources can be done through searching, which can be restricted by selecting relevant languages, media types, projects, or resource types. Many resources found in the repository have been developed in previous projects and the repository also contains links to externally hosted resources.

Digital Humanities Programme

SADiLaR's digital humanities programme aims to boost participation in the research field of digital humanities in South Africa. Initially, this was mainly done through the organization of training events, taught by researchers from SADiLaR's hub, in which interested participants (typically researchers and students) learned about computational tools and data collections. More recently, these training events have been restructured, allowing for a more fundamental approach where training events build on the learned skills and knowledge of previous events. In this context, a mentorship programme called Escalator² is currently under development.

The Escalator programme consists of several tracks, each targeting a different audience, from people interested in digital humanities to educators teaching digital humanities to their students. Each track has its own approach, from showcasing possibilities of existing tools, to hands-on collaboration and network building, to the development of educational material.

Challenges

One of the challenges related to the two programmes is how to best showcase computational tools and to provide insight in the availability and content of the digital language collections from the digitization programme. In this context, researchers at SADiLaR's hub investigate the usability of tools that show information from some of the language collections found in the repository. Note that preferably, these tools should

1. SADiLaR's repository can be found at: <https://repo.sadilar.org>.

2. Information on the Escalator programme can be found at: <https://escalator.sadilar.org>.

not require specific hardware and software to allow use by the general public as well.

Here, we describe initial investigations of tools that allow for investigation into geographically tagged language data, showing audio recordings on a map, illustrating different dialects spoken in the country. Below, two such approaches are discussed: Spraakatlas, which is used to place spoken (dialect) recordings on a map, and Glossopticon VR, which provides a virtual world that allows for “flying” over a map to listen to language recordings. Additionally, Byderhand, which provides access to location specific digital literature by scanning QR codes, is mentioned.

Case Studies

Spraakatlas

The Spraakatlas van Afrikaans (speech atlas of Afrikaans)³ is a website as shown in Fig 1, which visualizes dialectal differences of Afrikaans, one of South Africa’s official languages. By zooming in, it is possible to select a specific recording (see Fig 2) including some meta-information, such as location, gender, age, etc. By selecting the play button, an audio recording of Aesop’s fable “The North wind and the Sun ” can be heard in the dialect spoken at that location. This fable is selected because it contains all sounds used in Afrikaans.

Similar websites can be developed for other languages spoken in South Africa (e.g., based on existing spoken language collections in SADiLaR’s repository), allowing for fine-grained analyses of dialectal variation based on geographic location. For this to work, audio recordings of the different languages need to be available. Note, however, that the Spraakatlas already allows for the recording of speech (“Maak opname” translates to “Make recording”) as well as the required meta-data.

3. Spraakatlas van Afrikaans can be found at: <https://viva-afrikaans.org/lees-luister/spraakatlas>.

Spraakatlas van Afrikaans

Die vertaling van Esopus se fabel *Die noordewind en die son* wat vir hierdie projek gebruik word, bevat al die verskillende spraakklanke van Afrikaans.

Lulster: Klik op die vlaggies om na die groelende versameling voorlesings in verskillende Afrikaanse variëteite te luister en na die video's van die skrywer Anoeschka von Meck se reis deur Suid-Afrika – Op pad met Afrikaans – te kyk.

Neem op: Gee vir jou webblaaiër toestemming om jou ligging te bepaal. Klik op die swart mikrofoonikoon naby jou woonplek, lees die instruksies, en verwig ook jou tongval in Afrikaans se Spraakatlaskorpus. Die tweede opsie is om aan ons jou opname te stuur in die vorm van 'n stemboodskap ("voicenote") via WhatsApp.

Jy moet ingeteken wees om 'n opname te maak.

Maak opname

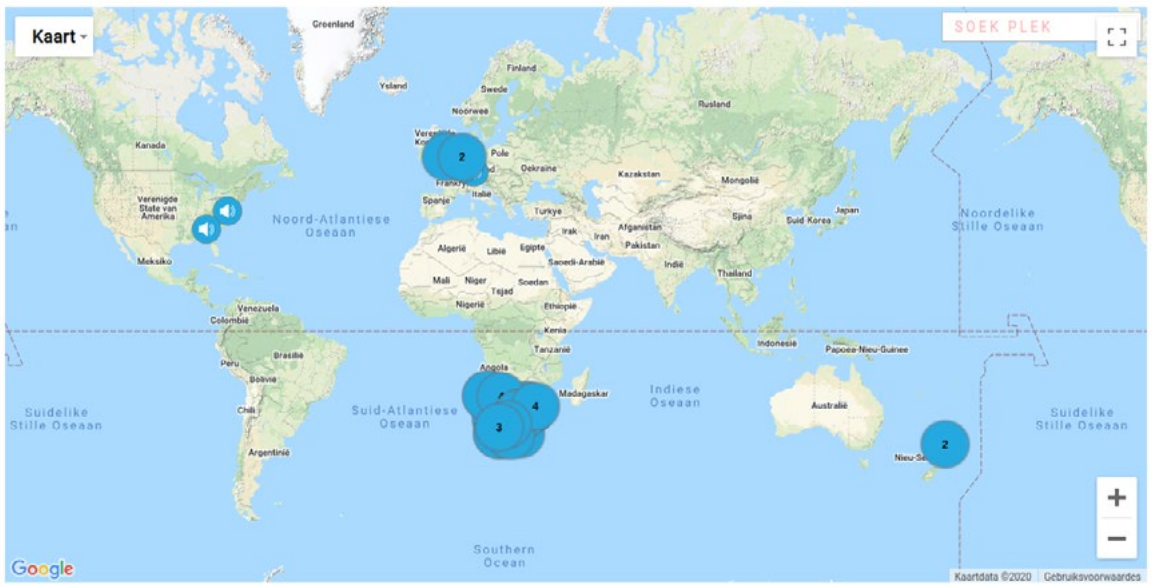


Figure 1. The main access page of the Spraakatlas van Afrikaans (speech atlas of Afrikaans).



Figure 2. Zoomed in image of the Spraakatlas van Afrikaans, showing meta-information on a single recording.

Glossopticon VR

The Glossopticon VR system⁴ (Burrell et al., 2019) offers language recordings and location information through virtual reality. This system currently contains over 1500 languages from the Pacific region. When starting the system, the user selects the relevant languages. Next, (part of) a map is visualized with vertical bars indicating the location of the different languages (see Fig 3). When “flying” through the bars, recordings of the respective language can be heard and some meta-information is displayed (see Fig 4). This approach is similar to the Spraakatlas system, but enhances interaction while retaining audio and location information.

4. Glossopticon VR can be found at: <https://glossopticon.com>.



Figure 3. Glossopticon VR interface illustrating the map as well as the vertical bars indicating the location of specific language recordings.

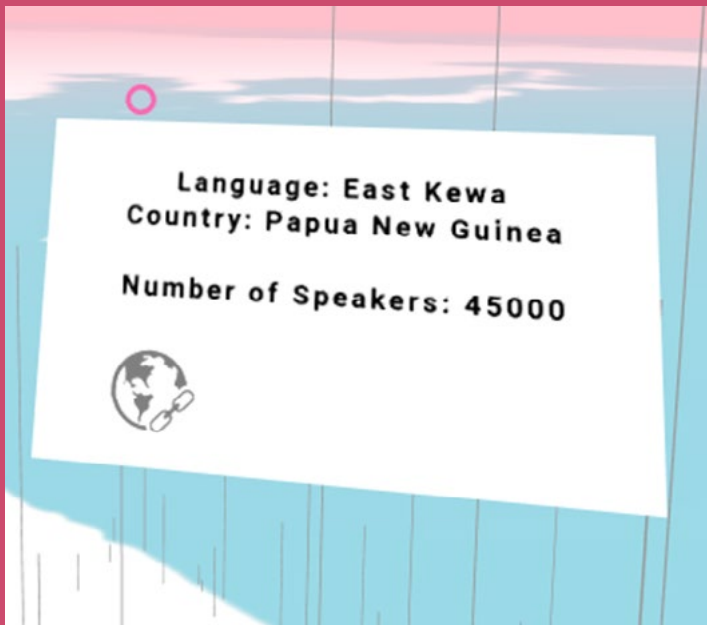


Figure 4. Glossopticon VR interface shows meta-information when a specific language is selected by hovering over a vertical bar at the same time, a recording of the language can be heard.

Byderhand

The Byderhand project⁵ (Odendaal, 2020; Greyling et al., 2020) allows for multimedia (e.g., text, audio, visual) documents to be placed in a real, physical environment. Using QR codes (as shown in Fig 5), one can access documents, such as digital literature, relevant to that specific location. The idea behind this project is to provide a digital form similar to rock art or graffiti. Note that this system is different from the previously discussed two cases in that digital information is placed in a real world environment, whereas the other two cases provide access to the language information in virtual environments.

5. Byderhand can be found at: <http://www.byderhand.net>.



Figure 5. On the left is an information plate with a QR code which provides access to the corresponding Byderhand multimedia document relevant to the specific location.

Discussion

The overview of the different systems described above is the result of initial investigations on how to tackle the problem described in the challenges section: to provide insight into existing computational language collections and at the same time to showcase (with virtual and augmented reality) digital humanities applications.

The Spraakatlas and Glossopticon VR systems can provide overviews of the content of speech related computational language collections. Without such systems, it is difficult for users to get a good grasp of what can be found in computational language collections. Obviously, such collections have meta-data attached, which provides a general overview and users can investigate the content on a sample basis. By placing the data points in a virtual space, it allows users to investigate the content of the data collection in a much more intuitive way.

At the same time, these systems (including Byderhand) provide interesting examples of how computational tools can provide insight as digital humanities tools. In general, it is not particularly easy to illustrate (e.g., to the general public) what digital humanities applications do if they do not have attractive interfaces. For instance, many computational linguistic and digital humanities experiments do not provide easily interpretable interfaces as the actual work typically takes place behind the scenes. The output of these experiments is interesting to the researcher, but even then may be difficult to appreciate by the general public. However, virtual and augmented reality systems (especially those described here) are directly understandable and provide a clear sense of usefulness.

Conclusions

SADiLaR aims to make available digital language resources (including language collections and computational linguistic tools) for each of the eleven official South African languages. Additionally, it aims to boost participation in the field of digital humanities among others through training events. One specific problem is that for many researchers it is unclear what language collections and computational linguistic tools are available and how these can be used.

Currently, several virtual and augmented reality approaches are investigated with two purposes in mind. Firstly, such systems allow for interactive investigations into large language collections. Secondly, these systems provide examples of practical digital humanities systems. The three cases discussed here provide examples on how digital tools may help provide insight into the linguistic data collections, or (for the third case) allow for novel ways of interacting with digital language data. Additionally, the different systems provide example applications that can be created using already resources. As for all discussed cases no special equipment is required existing, these are also accessible to the general public.

References

- Burrell, A., Hendery, R., & Thieberger, N. (2019, July). Glossopticon: visualising archival data. In *2019 23rd international conference in information visualization-part II* (pp. 100-103). IEEE.
- Greyling, F., Verhoef, S., & Tempelhoff, G. (2020). Die Byderhand-Pionierprojek:’n Gevallestudie van die deelnemende dinamika in die skep en toeganklikmaking van lokatiewe literatuur vir persone met siggestremdheid. *Tydskrif vir Geesteswetenskappe*, 60(4), 1336-1362.
- Odendaal, B. (2020). Manifestations of sensory perception in some poems included in two digital Byderhand installations at Worcester, South Africa. *Tydskrif vir Geesteswetenskappe*, 60(4-2), 1363-1385.
- SARIR (2016). South African Research Infrastructure Roadmap. 1st Department of Science and Technology.

Una visita a SADiLaR

Menno Van Zaanen

South African Centre for Digital Language Resources (SADiLaR)

menno.vanzaanen@nwu.ac.za

Resumen

El Centro Sudafricano de Recursos Lingüísticos Digitales (SADiLaR), financiado por el Departamento de Ciencia e Innovación de Sudáfrica, es una infraestructura de investigación sudafricana. La infraestructura de investigación consta de un hub y nodos en diferentes organizaciones (universidades, una organización de I + D y un centro interinstitucional) en todo el país. SADiLaR ejecuta dos programas: un programa de digitalización que se ocupa de la creación y distribución de recursos lingüísticos digitales (incluidas colecciones de idiomas escritos, hablados y multimedia, así como herramientas) para los once idiomas oficiales sudafricanos, y un programa de humanidades digitales que facilita la creación de capacidad de investigación en el campo de las humanidades digitales.

Aunque muchos investigadores en el campo de las humanidades y las ciencias sociales están interesados en las humanidades digitales, muchos no saben realmente por dónde empezar. Para intentar solucionarlo, SADiLaR organiza una serie de actividades, que recientemente se han estructurado dentro de un nuevo programa denominado Escalator. Una de estas actividades trata de mostrar investigaciones que muestran las posibilidades de las humanidades digitales de una manera accesible. Aquí, ilustramos investigaciones sobre sistemas que pueden usarse como vitrinas y, al mismo tiempo, pueden usarse para visualizar información que está contenida en colecciones de lenguaje digital (específicamente basadas en el habla) que también contienen información geográfica. SADiLaR considera el uso de varias herramientas, dos de las cuales se describen en este capítulo: Atlas de habla Africana y Glossopticon de Realidad Virtual (RV). Además, también se analiza una tercera herramienta, Byderhand, que proporciona acceso a documentos multimodales en

ubicaciones específicas del mundo físico. Actualmente, solo se han realizado investigaciones exploratorias sobre estas herramientas.

Palabras clave: Recursos del lenguaje digital, visualización, Spraakatlas, Glossopticon VR, Byderhand.

Introducción

El South African Centre for Digital Language Resources (SADiLaR) es un proyecto financiado por el gobierno dentro del programa South African Research Infrastructure Roadmap (SARIR) (SARIR, 2016). El centro consta de un centro, ubicado en la Universidad de North-West en Potchefstroom, Sudáfrica, y seis nodos: el Centro de Tecnología de Texto (CTextT) de la NWU que se ocupa de herramientas de lenguaje computacional, ICELDA (un centro interinstitucional) que se centra en las pruebas de idiomas, la Universidad de Pretoria trabaja en digitalización, la Universidad de Sudáfrica investiga terminología y Wordnets, el Consejo de Investigación Científica e Industrial (CSIR) actúa como nodo del habla y, recientemente, la Universidad de Stellenbosch se unió como nodo de desarrollo del lenguaje infantil. SADiLaR ejecuta dos programas, un programa de digitalización y un programa de humanidades digitales. Estos programas se apoyan a través de financiación (proporcionada por SADiLaR para proyectos en los diferentes nodos, pero SADiLaR también financia proyectos a través de convocatorias abiertas), así como apoyo práctico (por ejemplo, organización de eventos de formación, conferencias, etc.) por parte de investigadores en el centro. Estos programas tienen como objetivo proporcionar recursos que sean de acceso abierto, aunque en su mayoría son relevantes para estudiantes e investigadores en los amplios campos de las humanidades y las humanidades digitales.

Programa de digitalización

El programa de digitalización de SADiLaR desarrolla la creación sistemática de recursos lingüísticos digitales, incluida la creación de colecciones de texto, habla y multimodales para los once idiomas oficiales de Sudáfrica. Esto lo realiza, por ejemplo, el nodo de digitalización de SADiLaR en la

Universidad de Pretoria, aunque los proyectos adicionales que son financiados por SADiLaR también dan como resultado recopilaciones de datos de idiomas. Además, este programa desarrolla herramientas que permiten el análisis lingüístico computacional de estos lenguajes. Por ejemplo, el Center for Text Technology (CTeXt), uno de los nodos de SADiLaR, proporciona muchas de las herramientas disponibles.

Los diferentes recursos están disponibles a través de un repositorio¹. La identificación de recursos se puede realizar mediante la búsqueda, que se puede restringir seleccionando idiomas, tipos de medios, proyectos o tipos de recursos relevantes. Muchos de los recursos que se encuentran en el repositorio se han desarrollado en proyectos anteriores y el repositorio también contiene enlaces a recursos alojados externamente.

Programa de Humanidades Digitales

El programa de humanidades digitales de SADiLaR tiene como objetivo impulsar la participación en el campo de investigación de las humanidades digitales en Sudáfrica. Inicialmente, esto se hizo principalmente a través de la organización de eventos de capacitación, impartidos por investigadores del centro de SADiLaR, en los que los participantes interesados (generalmente investigadores y estudiantes) aprendieron sobre herramientas computacionales y recopilación de datos. Más recientemente, estos eventos de capacitación se han reestructurado, lo que permite un enfoque más fundamental en el que los eventos de capacitación se basan en las habilidades aprendidas y el conocimiento de eventos anteriores. En este contexto, se está desarrollando un programa de tutoría llamado Escalator².

El programa Escalator consta de varias pistas, cada una dirigida a un público diferente, desde personas interesadas en humanidades digitales hasta educadores que enseñan humanidades digitales a sus estudiantes. Cada pista tiene su propio enfoque, desde mostrar las posibilidades de las herramientas existentes, hasta la colaboración práctica y la creación de redes, hasta el desarrollo de material educativo.

1. El repositorio de SADiLaR se puede encontrar en: <https://repo.sadilar.org>.

2. Puede encontrar información sobre el programa Escalator en: <https://escalator.sadilar.org>.

Desafíos

Uno de los desafíos relacionados con los dos programas es cómo exhibir mejor las herramientas computacionales y proporcionar información sobre la disponibilidad y el contenido de las colecciones de idiomas digitales del programa de digitalización. En este contexto, los investigadores del centro de SADiLaR investigan la usabilidad de herramientas que muestran información de algunas de las colecciones de idiomas que se encuentran en el repositorio. Tenga en cuenta que, preferiblemente, estas herramientas no deberían requerir hardware y software específicos para permitir su uso también por parte del público en general.

Aquí, describimos las investigaciones iniciales de herramientas que permiten la investigación de datos de idiomas etiquetados geográficamente, mostrando grabaciones de audio en un mapa, ilustrando diferentes dialectos hablados en el país. A continuación, se analizan dos de estos enfoques: *Spraakatlas*, que se utiliza para colocar grabaciones habladas (dialecto) en un mapa, y *Glossopticon VR*, que proporciona un mundo virtual que permite “volar” sobre un mapa para escuchar grabaciones de idiomas. Además, se menciona a *Byderhand*, que brinda acceso a la literatura digital específica de la ubicación mediante el escaneo de códigos QR.

CASOS DE ESTUDIO

Atlas de habla Africana (El *Spraakatlas van Afrikaans*)

Atlas de habla africana (The *Spraakatlas van Afrikaans*³) es un sitio web como se muestra en la Figura 1, que visualiza las diferencias en los dialectos africanos, uno de los idiomas oficiales de Sudáfrica. Al hacer zoom, es posible seleccionar una grabación específica (ver la figura 2) que incluye cierta meta información, como ubicación, género, edad, etc. Al seleccionar el botón de reproducción, una grabación de audio de la fábula de Esopo “El viento del norte y the Sun” se puede escuchar en el dialecto que se habla en ese lugar. Esta fábula se selecciona porque contiene todos los sonidos africanos utilizados.

Se pueden desarrollar sitios web similares para otros idiomas hablados en Sudáfrica (por ejemplo, basados en colecciones de idiomas hablados

3. *Spraakatlas van Afrikaans* se puede encontrar en: <https://viva-afrikaans.org/lees-luister/spraakatlas>.

existentes en el repositorio de SADiLaR), lo que permite análisis detallados de la variación dialectal según la ubicación geográfica. Para que esto funcione, es necesario que estén disponibles grabaciones de audio de los diferentes idiomas. Sin embargo, tenga en cuenta que el Atlas de habla Africana ya permite la grabación de voz (“Maak opname” se traduce como “Hacer grabación”), así como los metadatos requeridos.

VIVA > LEES | LUISTER > SPRAAKATLAS

Spraakatlas van Afrikaans

Die vertaling van Esopus se fabel *Die noordewind en die son* wat vir hierdie projek gebruik word, bevat al die verskillende spraakklanke van Afrikaans.

LuiSTER: Klik op die vlaggies om na die groeiende versameling voorlesings in verskillende Afrikaanse variëteite te luiSTER en na die video's van die skrywer Anoeschka von Meck se reis deur Suid-Afrika – Op pad met Afrikaans – te kyk.

Neem op: Gee vir jou webblaaiër toestemming om jou ligging te bepaal. Klik op die swart mikrofoonikon naby jou woonplek, lees die instruksies, en verewig ook jou tongval in Afrikaans se Spraakatlaskorpus. Die tweede opsie is om aan ons jou opname te stuur in die vorm van 'n stemboodskap (“voicenote”) via WhatsApp.

Jy moet ingeteken wees om 'n opname te maak.

[Maak opname](#)

Figura 1. La página de acceso principal del Atlas de habla Africana (Spraakatlas van Afrikaans).



Figura 2. Imagen ampliada del Atlas de habla Africana (Spraakatlas van Afrikaans), que muestra meta información en una única grabación.

Glossopticon de Realidad Virtual (RV)

El sistema Glossopticon de Realidad Virtual (Glossopticon VR⁴) (Burrell et al., 2019) ofrece grabaciones de idiomas e información de ubicación a través de la realidad virtual. Este sistema contiene actualmente más de 1500 idiomas de la región del Pacífico. Al iniciar el sistema, el usuario selecciona los idiomas relevantes. A continuación, se visualiza (parte de) un mapa con barras verticales que indican la ubicación de los diferentes idiomas (ver la figura 3). Al "volar" a través de las barras, se pueden escuchar grabaciones del idioma respectivo y se muestra algo de meta información (ver la figura 4). Este enfoque es similar al sistema Atlas de habla Africana, pero mejora la interacción al tiempo que retiene la información de audio y ubicación.

4. Glossopticon VR se puede encontrar en: <https://glossopticon.com>.



Figura 3. Interfaz de realidad virtual de Glossopticon que ilustra el mapa y las barras verticales que indican la ubicación de las grabaciones de idiomas específicos.

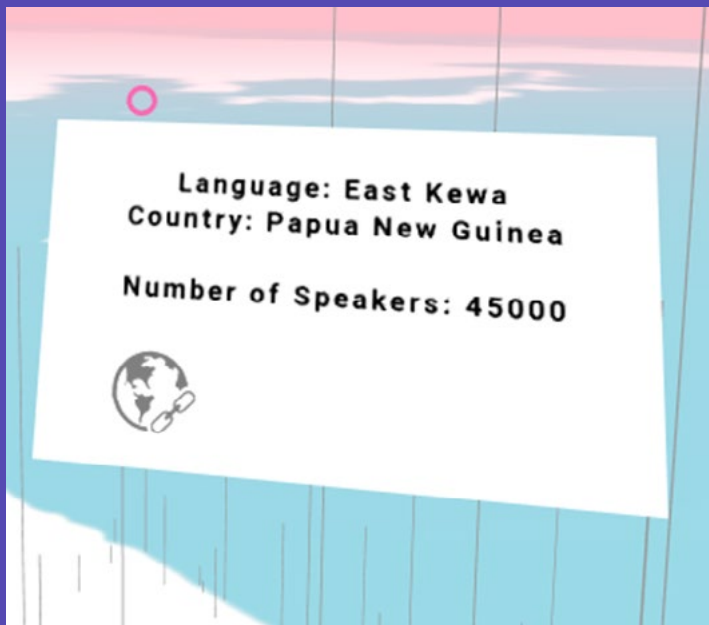


Figura 4. Interfaz de Glossopticon que muestra metainformación cuando se selecciona un idioma específico al pasar el cursor sobre una barra vertical al mismo tiempo, se puede escuchar una grabación del idioma.

Byderhand

El proyecto Byderhand⁵ (Odendaal, 2020; Greyling et al., 2020) permite colocar documentos multimedia (p. Ej., Texto, audio, visuales) en un entorno físico real. Usando códigos QR (como se muestra en la figura 5), uno puede acceder a documentos, como literatura digital, relevantes para esa ubicación específica. La idea detrás de este proyecto es proporcionar una forma digital similar al arte rupestre o al graffiti. Tenga en cuenta que este sistema es diferente de los dos casos discutidos anteriormente en que la información digital se coloca en un entorno del mundo real, mientras que los otros dos casos proporcionan acceso a la información del idioma en entornos virtuales.

5. Byderhand se puede encontrar en: <http://www.byderhand.net>.



Figura 5. A la izquierda hay una placa de información con un código QR que proporciona acceso al documento multimedia de Byderhand correspondiente relevante a la ubicación específica.

Discusión

La descripción general de los diferentes sistemas descritos anteriormente es el resultado de investigaciones iniciales sobre cómo abordar el problema descrito en la sección de desafíos: proporcionar información sobre las colecciones de lenguaje computacional existentes y, al mismo tiempo, mostrar (con realidad virtual y aumentada) las humanidades digitales.

Aplicaciones

Los sistemas de realidad virtual Atlas de habla Africana y Glossopticon pueden proporcionar descripciones generales del contenido de las colecciones de lenguaje computacional relacionadas con el habla. Sin estos sistemas, es difícil para los usuarios comprender bien lo que se puede encontrar en las colecciones de lenguajes computacionales. Obviamente, estas colecciones tienen metadatos adjuntos, que brindan una descripción general y los usuarios pueden investigar el contenido a modo de muestra. Al colocar los puntos de datos en un espacio virtual, permite a los usuarios investigar el contenido de la recopilación de datos de una manera mucho más intuitiva.

Al mismo tiempo, estos sistemas (incluido Byderhand) proporcionan ejemplos interesantes de cómo las herramientas computacionales pueden proporcionar conocimientos acerca de las humanidades digitales. En general, no es particularmente fácil ilustrar (por ejemplo, al público en general) qué hacen las aplicaciones de humanidades digitales si no tienen interfaces atractivas. Por ejemplo, muchos experimentos de humanidades digitales y lingüísticas computacionales no proporcionan interfaces fáciles de interpretar, ya que el trabajo real suele tener lugar entre bastidores. El resultado de estos experimentos es interesante para el investigador, pero incluso entonces puede ser difícil de apreciar por el público en general. Sin embargo, los sistemas de realidad virtual y aumentada (especialmente los descritos aquí) son directamente comprensibles y brindan un claro sentido de utilidad.

Conclusiones

SADiLaR tiene como objetivo poner a disposición recursos lingüísticos digitales, incluidas colecciones de idiomas y herramientas lingüísticas computacionales, para cada uno de los once idiomas oficiales de Sudáfrica. Adicionalmente, tiene como objetivo potenciar la participación en el campo de las humanidades digitales entre otros a través de eventos formativos. Un problema específico es que para muchos investigadores no está claro qué colecciones de idiomas y herramientas lingüísticas computacionales están disponibles y cómo se pueden utilizar.

Actualmente, se investigan varios enfoques de realidad virtual y aumentada con dos propósitos en mente. En primer lugar, estos sistemas permiten investigaciones interactivas en grandes colecciones de idiomas. En segundo lugar, estos sistemas proporcionan ejemplos de sistemas prácticos de las humanidades digitales. Los tres casos discutidos aquí proporcionan ejemplos sobre cómo las herramientas digitales pueden ayudar a proporcionar información sobre las colecciones de datos lingüísticos o permitir formas novedosas de interactuar con los datos del lenguaje digital. Además, los diferentes sistemas proporcionan aplicaciones de ejemplo que se pueden crear utilizando recursos ya existentes. Para todos los casos discutidos, no se requiere equipo especial y son accesibles para el público en general.

Referencias

- Burrell, A., Hendery, R., & Thieberger, N. (2019, July). Glossopticon: visualising archival data. In *2019 23rd international conference in information visualization-part II* (pp. 100-103). IEEE.
- Greyling, F., Verhoef, S., & Tempelhoff, G. (2020). Die Byderhand-Pionierprojek:’n Gevallestudie van die deelnemende dinamika in die skeep en toeganklikmaking van lokatiewe literatuur vir persone met siggestremdheid. *Tydskrif vir Geesteswetenskappe*, 60(4), 1336-1362.
- Odendaal, B. (2020). Manifestations of sensory perception in some poems included in two digital Byderhand installations at Worcester, South Africa. *Tydskrif vir Geesteswetenskappe*, 60(4-2), 1363-1385.
- SARIR (2016). South African Research Infrastructure Roadmap. 1st Department of Science and Technology.

MENNO VAN ZAAZEN



English

Menno van Zaanen was born in 's-Gravenhage, the Netherlands. He received an Msc in computer science from the Vrije University in Amsterdam, the Netherlands in 1997 and an MA in computational linguistics in 1998 from the University of Amsterdam, the Netherlands. Menno received his PhD in computer science from Leeds University, UK in 2002. He worked as research fellow at the University of Amsterdam, Tilburg University (both the Netherlands) and Macquarie University (Australia) as well as a guest researcher at the University of Groningen (the Netherlands). After that, he worked as a researcher, lecturer and assistant professor at Tilburg University, the Netherlands. Currently, he works as a professor in digital humanities at the South African Centre for Digital Language Resources (SADiLaR) which is hosted at North-West University, South Africa. His current research interests deal with unsupervised grammatical inference and applied machine learning to sequential data (in particular natural language). Among others, he worked on multi-modal structuring of data (music) and multi-modal information retrieval, as well on proofing tools (spelling checkers for African languages). More recently, he has been concentrating on applying digital techniques in the field of humanities. Menno is a founding member of the International Community in Grammatical Inference (ICGI), which he chaired from 2007-2010. Currently, he is chair of the Digital Humanities Association of Southern Africa (DHASA) and member of the Constituent Organization Board of the international Alliance of Digital Humanities Organizations (ADHO).

Español

Menno van Zaanen nació en la ciudad de 's-Gravenhage, Países Bajos. En 1997, Menno es magíster en ciencias de la computación de la Universidad de Vrije localizada en Amsterdam, Holanda. En 1998, obtuvo su segunda maestría como lingüística computacional en la Universidad de Amsterdam. En 2002, Menno obtuvo su doctorado en ciencias de la computación de la Universidad de Leeds, Reino Unido. Trabajó como investigador asociado en la Universidad de Amsterdam, la Universidad de Tilburgo (ambos Países Bajos) y la Universidad Macquarie (Australia). Fue investigador invitado en la Universidad de Groningen (Países Bajos). Luego trabajó como investigador, conferencista y profesor asistente en la Universidad de Tilburgo. Actualmente, trabaja como profesor de humanidades digitales en el Centro Sur Africano de Recursos Digitales en Lenguaje (SADiLaR) que se encuentra ubicado en la Universidad Norte Oeste de Sudáfrica. Sus intereses de investigación actuales se ocupan de la inferencia gramatical no supervisada y el aprendizaje automático aplicado a datos secuenciales (en particular, el lenguaje natural). Así mismo, trabajó en la estructuración multimodal de datos (música) y la recuperación de información multimodal, así como en herramientas de corrección (correctores ortográficos para idiomas africanos). Más recientemente, se ha concentrado en la aplicación de técnicas digitales en el campo de las humanidades. Es miembro fundador de la Comunidad Internacional de Inferencia Gramatical (ICGI), que presidió de 2007 a 2010. Actualmente, es presidente de la Asociación de Humanidades Digitales de África Austral (DHASA) y miembro del Consejo de Organización Constituyente de la Alianza Internacional de Organizaciones de Humanidades Digitales (ADHO).