Proceedings of the Inaugural Conference of the Digital Humanities Association of Southern Africa (DHASA) 2017

Stellenbosch, South Africa

17th- 20th January 2017

Keynote: Future Web: Co-shaping our Identities

Arthur, Paul
Australian National University

Virtual Keynote: Reconstructing Race: Languages of Identity in American Fiction, 1789-1964

Prof Mark Algee-Hewitt
Stanford University

Abstract

Digital humanities today is a vast, interdisciplinary field covering many different methods and understandings. Its object of study is just as often the distant past as it is contemporary culture. Increasingly engaged with computer science, digital humanities focuses on the development and use of digital tools and research methods to understand human experience and expression, especially through analysis, modeling, and manipulation of digital data. As new analytical possibilities have been developed, the focus in digital humanities has broadened. Data in the humanities can take many forms. Digital humanities asks fundamental questions about whether the analysis and manipulation of digital data can help to make greater sense of human society's cultural production and behavior. A contemporary theme running through digital humanities is how new knowledge practices are being opened up by 'big data' and data aggregation.

This talk focuses on biographical data, one of the most complex categories to address. Biography is just one of many disciplines that have been deeply influenced by advances in digital media and computing, and that have required new theoretical approaches to help understand the changes. Yet the digital revolution has arguably had a more profound effect on biography and life writing than on any other branch of literature, perhaps any branch of the arts. At the intersection of biography and digital humanities, key questions can be posed: In what ways does the Web act to co-shape our identities? Do we know ourselves, each other, or historical actors differently? How permanent are the digital records of lives that are being produced? Do we, or will we soon, remember differently? And, what are the research futures for digital biographical research?

Abstract

How does the discourse of race stabilize around different ethnic groups in different contexts in American long-form fiction? And how do these individual discourses change over time? As cultural objects, novels offer a unique point of contact with our social constructions of human racial and ethnic difference and the complex history of immigration in America has played out in the writing of American novels in ways that serve to both reinforce and resist specific configurations of identity. Drawing on 18,000 works of American Fiction (in the new Gale American Fiction bibliography), this project seeks to probe the language of race and ethnicity in American literature over two centuries, from Washington's inaguaration in 1789 to the passage of key parts the Civil Rights act in 1964. Through a detailed investigation into the specific configuration of language surrounding terms of identity, as well as novels that take various ethnic identities as their subject, this project explores the tension between the background discourse of race (the racial unconscious) and the ways in which individual characters react against (or with) the generalized language of ethnicity or ancestry.

Keynote: Imagining Ada Lovelace: creative computing and experimental humanities

Willcox, Pip
Bodleian Libraries, University of Oxford

Abstract

In the 200 years since Ada Lovelace's birth, she has been celebrated, neglected, and taken up as a symbol for any number of causes and ideas. A symposium to mark the 200th anniversary of her birth narrated many of these, including accounts of her generative relationship with Charles Babbage and his Difference and Analytical Engines. This talk traces some of the paths the idea of Lovelace has taken, what basis they have in her life, and what her reception tells us about our own scholarship and society. It goes on to describe our experimental work responding to Lovelace and Babbage, and to the operatic 'Ada sketches' of composer Emily Howard. We created a Web application to produce music from maths through programming a digital simulation of the Analytical Engine, after Lovelace's idea that ?the engine might compose elaborate and scientific pieces of music of any degree of complexity or extent?

The use of ATLAS.ti in a data analysis of Afrikaans speaking children's first form-meaning mappings

Brink, Nina
North-West University

Abstract

The aim of this presentation is to demonstrate how the computer software ATLAS.ti can efficiently be applied in the facilitation of an extensive linguistic study and as enhancement of the data analysis. This particular study would have taken up much more time and effort without the use of digital tools.

In Language Acquisition across Linguistic and Cognitive Systems Lieven (2010:92;103) indicates a universal need for more language acquisition studies in various languages and especially research on how form-meaning mappings are construed by children based on input, particularly in language specific contexts. The term form-meaning mapping refers to the way in which children pair certain meanings with certain word forms (lexical items). This process suggests a cognitive aspect, considering that a child has to conceptualise and categorise concepts in order to establish associations with a certain lexical item. In Language and its structure Langacker (1968:24) confirms that word-meaning pairings differ from language to language, except in the cases of words borrowed from other languages.

The focus of this cognitive semantic study aims to provide a grammatical description of the form-meaning mappings of Afrikaans children. A literature review suggests that no linguistic research has been done on this topic for Afrikaans children at the age at which they begin to acquire their first lexical items. Another research void in studies of Afrikaans language acquisition is the availability of original linguistic data. Therefore, data from 21 Afrikaans speaking children between the ages of 8 and 24 months using their first lexical items were analysed in this study in order to determine the types of form-meaning mappings these children make.

Theoretical assumptions from the usage-based theory and cognitive linguistics were used in the description of the language data of the children. The theoretical framework chosen from the usage-based theory in order to optimally describe the children's form-meaning mappings consists of cognitive models and mechanisms such as the prototype model of categorisation, metonymic and metaphoric mapping and image schemas. The process of conceptual blending is also used in describing the way in which children connect form and meaning.

The availability of computer technology greatly assists this line of research, since it facilitates access to potential respondents and enables the researcher to collect data accurately without travel. In the first place, the computer provided easy and convenient access to the respondents seeing that they, who live in various towns and

provinces, were contacted by means of e-mail and Facebook. It also made it easier for them to be willing to take part in the study because the method for data gathering was electronic diary entries (in Microsoft Word format) which were easy to complete and keep up to date for the five months during which the data were gathered.

The computer software for qualitative analysis, ATLAS.ti, was used for the coding of the data. Friese (2012:1) explains in Qualitative Data Analysis with ATLAS.ti that ATLAS.ti is very useful for the adjustment of codes and coded segments, the tracing of data on the grounds of certain criteria, the search for words, to integrate material, to attach notes to certain segments, to determine the number of codes as well as how many times a specific code is used, etc.

The electronic diary entries contain the lexical item used, three contexts in which the item is used and what the parent suspects the child means by the lexical item. The diary entries were loaded onto ATLAS.ti as primary documents in one hermeneutic unit. The three contexts of the lexical items were coded according to the kind of process taking place and the theoretical framework of the study. A number of rounds of coding took place after which the codes were further organised into code families. The use of ATLAS.ti made the coding process much faster seeing that 762 lexical items were coded.

Phase one of coding entailed that the lexical items were coded according to four different categories, namely unusable, non-symbolic imitation, taken over from adult language and complex mapping (further description necessary). These codes refer to the kind of form-meaning mapping that each lexical item is. During the second phase of coding the focus was on the lexical items coded as complex mapping in order to refine the coding of items according to the cognitive models and mechanisms as set out in the theoretical framework. Through the use of ATLAS.ti, going back to only these lexical items was easy and efficient. The in-depth analyses during this phase lead to the discovery that a number of different kinds of metonymic mappings can be distinguished. Examples of these metonymic mappings are mappings made

between a concept and a certain characteristic of that concept, e.g. the lexical item brrmm-brrmm (the sound a vehicle makes) used to include all types of vehicles such as cars, trucks, bicycles, etc. Another example of a metonymic mapping is the use of the lexical item pap 'porridge' to include any type of food. The frequency of metonymic mappings in the children's language indicates that this may be an active cognitive mechanism at the early stages of language acquisition. In this presentation, the types of metonymic mappings will be presented, as well as the way it can be structured into categories according to the prototype model of categorisation.

Furthermore, examples of the structuring of concepts according to image schemas are the use of klaar 'finished' to refer to an empty container (source-path-goal schema); the lexical item binne 'inside' to refer to inside, outside and when something is put into something else (container-schema) and the lexical item tak 'branch' when referring to a tree (part- whole schema). The structuring of experiences according to image schemas underlies the process of metonymic mapping. A further important finding is that metaphoric mapping is not prominent at all during this stage of language acquisition.

Therefore, the use of computer technology made the data very accessible and aided in the systematic approach and careful analysis of the data in order to come to certain conclusions.

The online adaptation of A Dictionary of South African English on Historical Principles.

Van Niekerk, Tim; Le Du, Bridgitte
Dictionary Unit for South African English

Abstract

The online Dictionary of South African English (DSAE, http://dsae.co.za) is an electronic version of A Dictionary of South African English on Historical Principles (Silva et al., Oxford University Press, 1996). The print edition, produced by the Dictionary Unit for

South African English (DSAE) in Grahamstown, South Africa was the culmination of 25 years' research resulting in a 1.7 million-word text with 4600 main entries documenting the development of South African English from its origins in the late 17th Century to 1995. Entries emphasise word history and show etymologies, variant spellings, compounds, derivatives and phrases. In total 14 700 word forms are represented, reflecting diverse borrowings from other South African languages; notably, the dictionary is rooted in quotation evidence, reproducing 44 000 bibliographically-documented citations.

In July 2014 a pilot online version of the dictionary was published (http://dsae.co.za) with initial funding from the DSAE's host institution, Rhodes University, to make this reference work available since it had been out of print since 2005. At the same time, a free access model was adopted. This was a first step towards a thorough, publicly-available adaptation of the print edition for electronic platforms. In 2015 the DSAE was joined by the University of Hildesheim, Germany and subsequently by the University of Stellenbosch, South Africa in an ongoing collaborative project to improve the initial pilot online edition to meet the needs of modern digital users. The target user group of the print edition was relatively specialised, namely linguists, historians, editors, writers, translators and interested laypersons. This target group has been broadened as part of the electronic adaptation process, to include a new audience of non-specialists. Given the length and complexity of some entries, this presents challenges not faced by synchronic dictionaries (e.g. trek, n., first used in English contexts in the early 19th Century and subsequently assimilated into General English, lists 11 main senses and is over 5000 words long; some exceptional entries are longer).

The presentation will give a description of this project, highlighting some key areas of the print-to-digital adaptation in the context of the evolving field of electronic lexicography, both at the level of entry layout components (article microstructure) and navigation features (access structure). To a large extent basic expected features of electronic reference works, as well as possibilities for innovative (e.g. visual) strategies for content display, depend on the internal computational representation of the dictionary text. Some examples will be given of how generic or unstructured representations of data categories in the 2014 dictionary dataset (encoded in XML) have had to be refined so that they are increasingly fine-grained in structure, to allow greater flexibility in presentation and to facilitate improved navigation features. Additionally, current work includes enriching the dataset with new metadata to transform the dictionary from a static, text-heavy reference work to a database-like tool supporting content filtering and browsing possibilities, for example through subject-categorisation of the c. 7400 senses.

Adaptive and optional presentation devices are being introduced, allowing users to 'show-more' or 'show-less' depending on their needs, and adding the possibility of exploring horizontal relationships between entries, not traditionally part of print dictionary structure beyond cross-references, via browsing and selective query or filtering functionality. The depth and range of content encoded in this dictionary, combined with the possibilities presented by the digital environment, thus allow this reference work to be transformed from the traditional concept of a dictionary as an extended wordlist to a linguistic, cultural and encyclopaedic inventory. Additionally, offering easy access to multidimensional lexicographical data through processes of print-to-digital adaptation requires the corresponding adaptation and supplementation of the 'front-matter' and 'end-matter' produced for the print edition with sometimes new types of outer texts, namely more context-sensitive and less text-heavy user guidelines, bibliography records, further information about the South African English variety (supported by infographics), and so forth. Past and current work on adapting and supplementing these texts to meet new requirements will be illustrated together with associated topics.

While the primary aim of the talk is to present current work and invite feedback on what is essentially an electronic dictionary publishing project, a further motivation is to draw attention to a structured dataset which may be of wider interest not only to lexicographers but also to computational linguists and Digital Humanities researchers. Although it is a lexicographical database, its diachronic, bibliographically-annotated design gives it historical, cultural and to some extent literary dimensions across a wide historical span. Until now, comparable datasets describing the South African variety of English and the multilingual influences acting on it have not been available and we hope that, parallel to the DSAE

dictionary project, these data may also help optimise collaborative knowledge generation across a number of overlapping fields.

African Wordnets: first steps towards use in the Digital Humanities

Bosch, Sonja E; Griesel, Marissa
University of South Africa

Abstract

The African Wordnet Project (AWN) aims at building wordnets for five African languages, namely Setswana, isiXhosa, isiZulu, Sesotho sa Leboa (Sepedi) and Tshivenda (see Griesel & Bosch, 2014 for a detailed introduction). The African wordnets are currently being developed manually by means of the expand model (Vossen, 1998) and are based on the English Princeton WordNet (PWN) (Fellbaum, 1998). A quality assurance phase formed part of the development of the roughly 55,000 synsets that are now included in the AWN.

Developing rich and hierarchically structured wordnets such as these for under resourced languages is labour intensive and costly. Apart from the various linguistic and lexicographic considerations that have to be made when developing a single entry in the wordnet, expert linguists also have to keep the interconnectivity of the meanings in mind. Synsets are linked in the different languages and to the English PWN in the structure of the database, but synonyms are also grouped together in different domains. This makes a wordnet in general and the AWN in particular useful for semantic analysis of texts in the African languages.

Any kind of annotation that adds information about the meaning of constituents in a text is regarded as semantic annotation. Semantically annotated or tagged corpora are considered one of the building blocks for natural language processing applications as well as investigations in the Digital Humanities (cf. Ide, 2004). These corpora are collections of texts that have been enriched with meanings in the form of synonyms for the key words. A definition, usage example and domain might accompany the tag. The resulting corpus can then be used in advanced (computerised) analysis of a text, machine translation or information retrieval. It is well-known that in languages such as English, wordnets have for many years served as a basis for sense tagging, one of the main reasons being that the Princeton WordNet for instance, is a freely available, online, machine-tractable lexicon providing extensive coverage of English.

At the moment no such resource exists for any of the African languages of South Africa, and therefore the development of the AWN can be considered as a first step towards the semantic analysis of texts and the subsequent creation of semantically tagged corpora. In this presentation we will show a novel application of wordnets in the Digital Humanities for the South African environment to create semantically tagged resources. The unique advantages of using a multilingual wordnet such as the AWN to gain semantic access to texts as an L2 reader will also be illustrated. We will conclude with recommendations for future development.

References:

Fellbaum, C., (ed), 1998. Wordnet: An electronic lexical database. The MIT Press, Cambridge, Mass. ISBN 978-0262- 06-197-1

Griesel, M. & Bosch, S. 2014. Taking stock of the African Wordnet Project: 5 years of development. In Orav, H., Fellbaum, C. &

Vossen, P. (eds): Proceedings of the 7th Global WordNet Conference 2014 (GWC2014), pp. 148-153, Tartu, Estonia, ISBN 978-9949-32-492-7.

Ide, N. 2004. Preparation and Analysis of Linguistic Corpora. In Schreibman, S., Siemens, R. & Unsworth, J. (eds): A companion to Digital Humanities. Chapter 21. Oxford: Blackwell. [online]
http://www.digitalhumanities.org/companion/. Accessed on 21 Octoboer 2016

Vossen, P. 1998. EuroWordNet: A multilingual database with lexical semantic networks. Kluwer Academic, Dordrecht. ISBN 0-7923-5295-5

Can Digital Approaches Serve as Catalyst for Harvesting Indigenous Knowledge?

Alexander, Josephine Olufunmilayo
University of South Africa

Abstract

At the First Digital Humanities Workshop held at North West University, Potchefstroom in 2015, illuminating accounts were given of the work done in Europe and elsewhere in the digital humanities field. The potential for reinventing the way we conduct research in the humanities and social sciences, through such interdisciplinary approaches, became clear. This raised the possibility of how digital humanities can advance the cause of intellectualizing African indigenous knowledge. In particular, how might digital humanities harvest the knowledge embedded in indigenous systems and traditional practices which are being eroded because of their non-transmission from one generation to another? This paper is also conceived against the background of the current debate on decolonising and Africanising education. Using the language and cultural traditional practices of the Owé people of Nigeria as a springboard, I attempt to explore the question, to provoke debate and to guide future research that could both benefit the Owé community and advance the cause of scholarship in the digital humanities field in an African context. The paper starts with an exploration of digital humanities and digital approaches. This is followed by an exposition of Owé language and the cultural and traditional practices embedded in it. Thereafter, an attempt is made to conceptualise how the approaches could serve as restorative agents for documenting, maintaining and preserving Owé indigenous knowledge systems and by extension other endangered African languages. The paper concludes by highlighting the scholarly and intellectual benefits of harnessing indigenous knowledge through the interrogation of digital approaches.

Utilising emerging technologies to design an immersive artistic experience to create awareness about human trafficking.

Steyn, Juan; Fourie, Heniel
North-West University

Abstract

Context

This paper reports on work being done as part of Disappearance, the fourth interdisciplinary practice-led research project conceptualised within the Research niche Visual Narratives and Creative Outputs through Interdisciplinary and Practice-led Research (ViNCO) at the North-West University, Potchefstroom campus (Chan & Pretorius, 2016).

The project's roots lie in the relationships between man and nature and the respective power position of man in this relationship. Using the theory of anthropocentrism, that sees mankind as the central and most significant entities in the world (Boslaugh, 2016), as a starting point, the project aims to explore how this relationship plays out within the context of poaching and human trafficking.

This paper will only focus on the design of an immersive artistic experience to create awareness about human trafficking as a sub-project within the larger Disappearance project.

Immersive artistic design and human trafficking

Designing interactive or immersive spaces utilising technology is not something new (Mitchell, 2010), however within the domain of digital humanities internationally the use of emerging technologies applied to humanities research is gaining popularity as seen in the recent call for papers for the 2017 Alliance of Digital Humanities Organisations conference.

The authors decided to design an artistic installation that enables the spectator to become directly involve and interact with the installation itself. This is done in contrast to typical static descriptions that one will find on websites creating awareness about human trafficking. The aim is to give the participant an experiential understanding concerning human trafficking rather than just a static overview.

In order to construct an installation that can deliver the described experience, the space it occupies has to be activated. Creating an activated space through the use of technology has its challenges when using conventional technology infrastructure. Therefore, in order to minimise the footprint of the technology, single board computers will be used to power the multisensory experience of the installation.

Aspects that is needed to create an immersive experience include:

1. Motion sensors to track participant movement within the installation.

2. Static sensors that allow participants to engage with a specific part of the larger installation.

3. Server functionality to deliver content on an on-demand basis to increase the level interactive experience

   within the installation.

4. A narrative that would guide the participant through the installation.

The planned installation will involve a constructed room that has interactive stations and sections integrated into its design. Exploiting the computing power and aspects described above, a narrative will be played out in the form of an on-demand short film. This film will aim to immerse the onlooker in a human trafficking narrative. The multisensory environment will be further enhanced through the use of location-specific ambience and tactile experience.

The paper will also discuss the importance of interdisciplinary collaboration within the humanities to create quality digital humanities projects.

References

Boslaugh, S. 2016. Anthropocentrism https://global.britannica.com/topic/anthropocentrism Chan, J. & Pretorius, F. 2016 Disappearance

Mitchell, B. 2010. THE IMMERSIVE ARTISTIC EXPERIENCE AND THE EXPLOITATION OF SPACE

http://www.bcs.org/upload/pdf/ewic_ca10_s3paper2.pdf

Creative exploration in literature, place and technology.

Tempelhoff, Wilhelm Gustaf; Steyn, Juan
North-West University

Abstract

Faster and more efficient networks, platforms and mobile devices have enabled us to explore interesting concepts like site specific digital art and literature. Site specific digital literature is texts that are place bound and accessed via mobile devices. An example of this is Byderhand (Byderhand.net, 2016), a site specific digital literature project that was launched during the 2015 Aardklop National Arts Festival in Potchefstroom. The project incorporated the use of QR- codes, which enabled users to gain access to multimodal texts on their mobile devices. Byderhand was awarded the Aartvark prize for a ground breaking production, and the team was approached to host similar productions elsewhere. In 2016 the Byderhand project and platform was used in various educational

contexts. Most recently 1st additional language (Afrikaans) learners at HTS Potchefstroom were given the opportunity to publish their own site specific digital literature on the school grounds with the assistance from the Byderhand-team.

The possibilities of site specific digital literature opens doors for creative expression, place making and new multimodal user experience of texts in a non-traditional way. Site specific digital literature can also contribute to a better understanding of the relationship between digital interfaces, spatiality and people (Kaye, 2000:203; Turner & Davenport, 217-220)

Although there is a demand for expansion of the Byderhand-project there are some limiting factors a team of dedicated members is required to develop and implement such a project. This is costly and not scalable. Furthermore, infrastructure is needed in the form of servers and physical QR-codes that need to be maintained and updated to ensure functionality. On the user end, mobile data cost is also an constraining factor.

However, it is possible to develop an automated platform and basic infrastructure that can empower a user with very basic skills, to create and publish their own projects. Infrastructure can be provided by a single board computer such as the Raspberry Pi and the platform can be based on open standards.

This paper reports on the prototyping of such a system and considers the following:

1. Can the prototype platform enable users to publish multimodal texts and multimedia (e.g. audio, images and video) on their own.

2. Can the prototype platform provide an easy way to consume published work using entry level smartphones.

3. What possibilities does this hold for the expansion of multimodal site specific literuature and the building of of site spesific corpora.

4. What educational possibilities exist that can introduce students to the world of digital literature and the field of Digital Humanities.

References

byderhand.net 2016. interaktiewe Leeservaring http://byderhand.net/

Kaye, N. 2000. Site-Specific Art: Performance, Place and Documentation

Turner, P. & Davenport, E. eds. 2006. Spaces, Spatiality and Technology

An introduction of ISO/TC37: Terminology and other language and content resources

Janke, Ulrike[1]; Bosch, Sonja[2]; Calteaux, Karen[3]
Department of Arts and Culture[1]
University of South Africa[2]
Council for Scientific and Industrial Research[3]

Abstract

This presentation provides an overview of the work of the Technical Committee 37 of the International Standards Organisation. The standards it produces are applicable to researchers and developers working in the wider language industry and anyone involved in information and knowledge creation and management. Some relevant standards that have been published will be introduced, as well as those that have been adopted by South Africa through the South African Bureau of Standards (SABS). Current work of the committee will also be presented.

ISO/TC37 serves all fields and applications where human-human and human-machine communication is involved. This refers in particular to the language industry, in the development of terminology, human language technology applications and other language resources, as well as language services such as translation and interpretation. It further refers to fields and applications where knowledge is represented in whatever form. Thus ISO/TC37 standards are fundamental for language resource management, knowledge management and content management.

ISO/TC37 was originally established in order to standardise the methodology of terminology standardisation, and over time its scope grew to

cover language and content resources and applications in any field or application which involves communication as well as the creation, management and transfer of information and knowledge. Its scope includes the standardisation of principles, methods and applications relating to terminology and other language and content resources in the contexts of multilingual communication and cultural diversity. ISO/TC37 consists of five subcommittees: (1) Principles and methods, (2) terminographical and lexicographical working methods, (3) systems to manage terminology, knowledge and content, (4) language resource management, and (5) translation, interpreting and related technology.

The objective of ISO/TC37 is to prepare standards specifying principles, methods and data structures for creating, managing, coding and processing language resources. Its mission is to provide standards and guidelines to standardisation experts and language professionals who are involved in creating and handling language resources.

Users and target groups of ISO/TC37 include language-related professionals (translators, interpreters, technical writers, language teachers, linguists, lexicographers, terminologists), language planners and policy makers, and developers of IT- related software, such as computer-assisted translation, machine translation, content management, automatic classification and knowledge management, cataloguing, and indexing.

Content creation and management as well as system design for e-business rely on ISO/TC37 standards to ensure interoperability, reusability and machine processability of language data. The methodologies laid down in ISO/TC37 documents facilitate the development of more efficient and effective modes of access to information. These standards lead to the development of language resources, such as terminology databases and text annotation schemes, which are multipurpose and interoperable. They also support innovation and development in natural language processing technologies and language engineering, which are critical to the future evolution of computing.

Worldwide application of ISO/TC37 standards helps to enhance the overall quality of terminologies and other language resources in all subject fields, to improve information management within various industrial, technical, commercial and scientific environments, to reduce its costs, and to increase efficiency in technical standardisation and professional communication.

A stylometric analysis of Joseph Conrad's writing

Botha, Lande; Van Zyl, Maryka; Pienaar, Wikus
North-West University

Abstract

The language and style in the writings of Joseph Conrad, a multilingual, non-native speaker of English, have been the topic of various studies (Monod, 2005; Peters, 2006; Ophir, 2014; Simmons, 2014). Dowden (1973), Lucas (1991), Stubbs (2005), Moon (2007), Nofal, (2013), Hunter and Smith (2014) have paid a great deal of attention to different aspects of grammar in Conrad's writing in order to describe his idiosyncratic style. These studies mostly focus on analysing and describing solitary or a small handful of texts in terms of their socio-cultural, political and personal affect. Digital versions of Conrad□s works and digital analysis tools now make it possible to conduct a quantitative study of Conrad□s linguistic style in all of his writing (both fiction and non-fiction) spanning his working life. The aim of this study is to statistically compare texts from various genres (novels, novella, autobiography and notes) and publication times in order to establish whether a consistent "Conradian" style is maintained across genres and time. A "corpus" of all the published works of Conrad serves as input for Stylo (0.6.0), an R-script (Eder & Rybicki, 2011). Stylo provides a cluster tree analysis in which each text is positioned according to its relation to (stylistic distance from) every other text. This analysis is based on the hundred most frequent words and gives an indication of the extent to which genre and time are factors in the lexical choices of Conrad. The wordlist and keywords function in

WordSmith Tools (6.0) (Scott, 2012) allow for further lexis-based comparison of the texts. For purposes of a keywords analysis, the texts are grouped into two (or three) corpora based on the first branching in the Stylo-generated cluster tree. It is also possible to move beyond lexis and to study the grammatical aspects of Conrad□s style quantitatively by making use of a part-of-speech-tagged version of the corpus. CLAWS4 (Garside & Smith, 1997) is used to tag the data. A (Pearl) script strips the words from the POS-tags leaving "texts" consisting entirely of word class designations. These texts serve as input for a cluster tree analysis in Stylo using bigrams, and then trigrams, which allow for comparison of the texts based on grammatical structure. This indicates the extent to which genre and time of publication are factors in the grammatical choices made by Conrad. The POS-"texts" can also be grouped into two (or three) "corpora" based on the first branching in the cluster tree analysis to serve as input for a "keywords" analysis in Wordsmith. Such an analysis gives an indication of the word classes involved in grammatical differences between the texts. Most of the CLAWS tags also contain morphological information such as tense, aspect and number giving a richer picture of the author□s style.

References

Dowden, W.S. 1973. Joseph Conrad: the imaged style. Style, 7(2):245-248.

Eder, M. & Rybicki, J., 2011, Stylo version 0.6.0, computer software, Poland: University of Kraków.

Garside, R. & Smith, N., 1997, „A hybrid grammatical tagger: CLAWS4□, (In Garside, R., Leech, G. & McEnery, A., eds. Corpus annotation: Linguistic information from computer text corpora. London: Longman. p. 102-121).

Hunter, S. & Smith, S. 2014. A network text analysis of Conrad□s Heart of Darkness. English linguistics research, 3(2):39-53.

Lucas, M.A. 1991. Conrad□s adjectival eccentricity. Style, 25(1):123-150.

Monod, S. 2005. Joseph Conrad□s polyglot wordplay. Modern language review, 100:222-234.

Moon, R. 2007. Words, frequencies, and texts (particularly Conrad): a stratified approach. Journal of literary semantics, 36(1):1-33.

Nofal, K.H. 2013. Darkness in Conrad□s Heart of Darkness: a linguistic and stylistic analysis. The Buckingham journal of language and linguistics, 6:77-93.

Ophir, E. 2014. "All our stammerings": two kinds of inarticulateness in Conrad. A quarterly journal of short articles, notes, and reviews, 27(1):23-27.

Peters, J.G. 2006. The Cambridge introduction to Joseph Conrad. New York, USA: Cambridge University

Scott, M., 2012. WordSmith Tools version 6, computer software, Liverpool: Lexical analysis software.

Simmons, A.H. 2014. Reclaiming Conrad from his editors: the case of An Outcast of the Islands. Conradian, 46(1-2):39-52.

Stubbs, M. 2005. Conrad in the computer: examples of quantitative stylistic methods. Language and literature, 14(1):5-24.

Beyond Toponyms. Conceptualising space in narratives using networks and ontologies.

Viehhauser, Gabriel; Barth, Florian
University of Stuttgart

Abstract

Despite its importance as a fundamental category for the evolvement of narrative action, the conceptualisation of space has long been neglected in narratological research, since it poses substantial problems for modelling: The creation of space in narratives is often based on implicit information. Rather than constructing a given, mathematical space beforehand, stories tend to evolve their setting in relation to its characters that constitute space through their actions.

Besides, spatial information can also be used for descriptions that do not contribute to the creation of the setting of a story. Recent narratological studies thus have proposed to differentiate between event regions and mentioned spatial objects (Dennerlein 2009).

Given this preliminaries, an attempt to establish digital methods for the analysis of spatial information in narratives has to cope with complex semantic structures. Therefore, we propose a combination of a lexicological (1) with a relation extraction-approach (2), which we will outline in our paper.

(1) A basic approach to detect spatial information can be achieved by extracting toponyms from a text, which can also be visualised in a map. However, toponyms are not the only relevant spatial information in a text. Dealing with German texts, we use the GermaNet-ontology (Hamp/Feldweg 1997, Henrich/Hinrichs 2010) to build wordlists of place nouns that are divided into sub-categories (architecture, landscape, etc.) and combine them with the information on toponyms extracted by Named Entity Recognition.(2) To classify spatial information, it has to be seen in its context. Although a rule-based approach that draws on dependency parsing and the detection of verb-classes (verbs of motion vs. verbs of perception)

appears promising for separating event regions from mentioned spatial objects, in this paper we constrain ourselves on collocation- networks as they are used in the analysis of character relations in stage plays (e.g. Trilcke 2013). Instead of interacting characters we use place markers to serve as nodes of the network. Edges are established according to their common appearance in one sentence.

(2) To classify spatial information, it has to be seen in its context. Although a rule-based approach that draws on dependency parsing and the detection of verb-classes (verbs of motion vs. verbs of perception) appears promising for separating event regions from mentioned spatial objects, in this paper we constrain ourselves on collocation- networks as they are used in the analysis of character relations in stage plays (e.g. Trilcke 2013). Instead of interacting characters we use place markers to serve as nodes of the network. Edges are established according to their common appearance in one sentence.

An example of this approach can be seen in figure 1, which contains a network from chapter 14 of Jules Verne's Around the World in Eighty Days (place nouns have been translated into English). Toponyms that have been manually classified as event regions appear in brown, mentioned toponyms in red. Place nouns are divided into the sub-categories nature (green), architecture (grey) and transport (blue). The network not only visualises the relations between spatial markers, but can also serve as a starting point for testing hypothesis on their distribution (Do, for instance, mentioned spatial objects have a lower degree than event regions? Do the clusters of place markers indicate distinct settings of the story?).

However, this first approach towards a visualisation of space still has to overcome some difficulties: Firstly, even with the use of GermaNet, the wordlists have to be refined semi-automatically. Secondly, there is the problem of ambiguity (consider e.g. the terms "stream" or "railroad"), which has to be solved by more elaborated semantic analyses. Thirdly, the common appearance of place markers can only

serve as a proxy; here, more elaborated models of spatial relations have to be found.

References

Bastian Mathieu / Heymann Sebastian / Jacomy Mathieu (2009): Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media.

Dennerlein, Katrin (2009): Narratologie des Raumes. Berlin.

Hamp, Birgit / Feldweg, Helmut (1997): „GermaNet – a Lexical-Semantic Net for German.", in: Proceedings of the ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications. Madrid.

Henrich, Verena / Hinrichs, Erhard (2010): "GernEdiT – The GermaNet Editing Tool", in: Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC 2010). Valletta, Malta, May 2010 2228-2235.

Trilcke, Peer (2013): Social Network Analysis (SNA) als Methode einer textempirischen Literaturwissenschaft, in: Ajouri / Mellmann / Rauen (Ed.): Empirie in der Literaturwissenschaft, Münster, 201-247.

The role of computer-mediated analyses to improve validity of linguistic evidence for Forensic Linguistic purposes

van den Berg, Karien
North-West University

Abstract

The focus of this paper is the role of computer-mediated analyses to address concerns regarding the validity of linguistic data analysis for Forensic Linguistic purposes. Increasingly linguists are playing an active role in rendering court judgement in. for example, authorship attribution cases. Linguistic testimony, however, is not always accepted or admissible as evidence in a court of

law owing to, amongst other things, reliability concerns and the belief that no expertise is required for determining the meaning or intended meaning of a text (Tiersma and Slolan, 2002). Following the establishment of the Daubert standard in 2000, requirements for accepting expert interpretations in court have become more stringent and applies equally to linguistic evidence as to other scientific measures that inform expert opinion. The basic requirements of these standards entail that expert opinion be based on sufficient data, achieved through appropriate and valid application of reliable principles or theory and methods given the context of the case. This requires validity, accessed through a process of validation: collecting relevant linguistic evidence by appropriate means to support claims concerning legal issues. Computational developments provide objective means of achieving such evidence and may greatly enhance the validity claims offered by linguistic experts for legal scrutiny. With the aim of demonstrating how a validation argument for Forensic Linguistic evidence may be formulated and the importance of computer assistance (e.g. through web-scraping for ground truth data, Chaski:2013) to enhance validation procedures, this paper explores methods employed or proposed by leading forensic linguists MacMenamin (1993); Shuy (2008); and Chaski (2013). Conley and O'Bar (2005:177) advise the forensic linguist to follow a linguistically driven approach to the field, looking at the law from a linguistic view point, rather than being linguists working for the law. Three main approaches to forensic linguistic analysis are discernible: forensic computational linguistics, forensic stylistics and stylometric computing. Arguing in favour of the first of these, Chaski wishes similar rigour (based on validity tests) in applying stylistic and stylometric methods "so that reliable methods of forensic authorship identification can be offered to our courts" (2013:372). I would therefore like to present Chaski's claims in terms of a validation argument as propagated by Kane (2013) and Davies and Elder (2011), emphasising the role of computational methods in strengthening the evidence. This claim may then be compared to

that presented by Shuy and MacMenamin to determine the comparative strength of these arguments. In addition, three South African based case studies (Hubbard, 1994; Kotzé, 2007 and Grundlingh, 2015) are considered in comparison to establish the extent to which local practice may align with, or differ from international practice and standards, and adhere to validity requirements. Recommendations are made for using specific computerised tools to enhance evidence based on linguistic expertise in, for example, authorship attribution cases, to be useful and valuable in a court of law. Specific consideration is given to using computer tools in overcoming challenges typical to the field of authorship attribution, viz. very small data sets and calculating correct error rates.

That Hideous Strength (1945): CS Lewis and Charles Williams.

du Plessis-Hay, Michèle; Steyn, Juan
North-West University

Abstract

Background

That Hideous Strength (1945), the third novel in CS Lewis's Ransom Trilogy (1938–45), is frequently described as "a Charles Williams novel by CS Lewis" (Walter Hooper, CS Lewis: A Companion and Guide (1996), 231; Brian Horne, "A Peculiar Debt: The Influence of Charles Williams on C.S. Lewis" (1990), 93–95). Between 1930 and 1945, Charles Williams published seven novels: War in Heaven (1930); Many Dimensions (1930); The Place of the Lion (1931); The Greater Trumps (1932); Shadows of Ecstasy (1933); Descent into Hell (1937); All Hallows' Eve (1945). In each of these novels, the supernatural impinges — sometimes violently — on the ordinary phenomenological world, and characters undergo spiritual growth or change because of the supernatural incursions.

Lewis was undeniably influenced by Williams, whom he admired, but an easy equation of the two as authors may be questioned or modified. Lewis first read a novel by Williams in 1936: the first novel of his trilogy (Out of the Silent Planet) was published in 1938 and the second (Perelandra) in 1943, but these novels are not described as being "Williams novels by Lewis". The difference in reception may be because the first two novels are set on Mars and Venus, while That Hideous Strength has an earthly setting, like all seven of Williams' novels (Hooper 231; Horne 94). In addition (as Lewis argued) the combination of the marvellous, fantastic, or supernatural with the ordinary was not a new thing, and Lewis and Williams, in much of their fiction, were continuing an existing tradition (CS Lewis, "The Novels of Charles Williams" (1982), 46–7, 49).

There are certainly elements in That Hideous Strength that are similar to elements in Williams' novels. Some are particular acts of homage, others arise from shared interests, and others arise from the combination of supernatural and ordinary in the lives of the characters — and hence will share characteristics with all fiction of that kind. But this is still not enough either to praise or dismiss That Hideous Strength as "a Williams novel by Lewis".

This is as far as ordinary literary critical argument can take us. But maybe some of the techniques of digital humanities can add weight to one or the other side of the debate. Stylo, an R package for stylometic analysis, may be able to show whether That Hideous Strength is closer to the other two novels of Lewis's trilogy or to Williams' novels. (Of course, Williams' novels — published over sixteen years, but with the first five over four years — may prove not be stylistically homogenous: this would complicate and enrich the questions and answers that would follow.) And, whatever Stylo shows, exploration using desktop software such as Wordsmith Tools and/or web-based Voyant Tools would enable those showings to be explicated.

Preliminary findings

The consensus of various sampling methods in Stylo is that the first two novels in Lewis's trilogy cluster together, while That Hideous Strength clusters more closely with novels by Williams. These clusters vary with re-runs of random

samples, suggesting that different parts of the novel exhibit different levels of closeness to Williams' novels.

Preliminary word counts using Voyant Tools also show That Hideous Strength having more in common with Williams' novels, while the first two of Lewis' novels differ from all the other novels under consideration, and in similar ways. Word counts suggest that Out of the Silent Planet and Perelandra contain a somewhat smaller proportion of dialogue than the other novels. They also suggest that they contain a greater proportion of similes and metaphors. Questions:

The paper will conclude with the beginnings of answers to at least some of the following questions.

Granted that That Hideous Strength is stylistically closer to Williams' novels than Out of the Silent Planet and Perelandra, in what ways is this so? How does That Hideous Strength differ from the other novels in Lewis's trilogy? How is it similar to Williams' novels – and is it equally close to all of them? Do different parts of the novel (which moves between three distinct settings and groups of characters) have clearly different relations to the other novels under consideration? Does Stylo generate the same results for the novels if they are compared in terms of grammatical structure (using n-grams and part-of-speech tagging) rather than lexical content?

How do the novels compare in terms of the ratio of narrative/exposition to dialogue? How do they compare in terms of their use of comparative language such as similes, metaphors, and analogies (this question may be applied to dialogue, narrative/exposition, or both)? Wordsmith Tools, especially the "Tagging and Markup" tools, will facilitate the answering of these questions.

Once these questions are answered, it will be possible to ask again whether, or in what way, or to what degree, That Hideous Strength might be considered "a Charles Williams novel by CS Lewis". The discussion might then return to the domain of literary criticism — but greatly

enriched; and the analyses may generate hitherto unsuspected areas for exploration within digital humanities.

Orthography-based Language Modelling for Speech Recognition.

Sindana, Daniel; Manamela, Madimetja Jonas
University of Limpopo

Abstract

Automatic speech recognition (ASR) is a branch of natural language processing (NLP) that uses computer algorithms to convert speech data to its corresponding textual representation in the form of a phone/word sequence [1]. One of the major components of an ASR system, as shown on Figure 1, is a language modelling (LM) component that is concerned with creating models of phone/word sequences to capture regularities of a language such as syntactic, semantic and pragmatic characteristics [2] [3]. When developed, the language models are used by a speech recogniser to determine the likelihood of unknown word sequences to be valid or legal sentences of the language. The more knowledgeable these models are about the target language and/or the recognition task, the better their determination of the most likely phone/word sequences matching the utterances.

This study investigates the influence that the orthography (or conventional writing system) of a language has on the quality of language models created for speech recognition. Different language modelling approaches and/or recipes that take into account the effect that a unique language structure has on the performance of its language models, have been recommended for different languages. Unique language structures such as rich morphology, high inflexion, flexible word order, and compounding of words by different languages have been investigated [5] [6] [3] [7] [8] [9] [10] [11]. In this study we focus on the unique conjunctive and disjunctive writing systems of IsiNdebele (endonym for Ndebele) and Sepedi (endonym for Pedi) respectively, in relation to language modelling. Towards this goal,

different language models are created and analysed.

In this paper, we report the results from three preliminary language modelling experiments conducted using the Lwazi and NCHLT text corpora, on the SRILM toolkit. The language modelling results indicate the need to model the writing systems of the chosen languages differently.

Thus far, we have conducted only three of the planned ten experiments. For experiments 1 and 2, we used default parameters to develop the LM models up to n-gram order 6; and up to order 20 for experiment 3. We used LM smoothing methods supported by the toolkit in the experimentation process. In all the experiments, either the Ndebele or Pedi test data set were used for testing the developed models.

In Experiment 1, the text was normalised in different ways and resulted in the following versions: original tagged text without start and end sentence tags – e.g., <s> and </s>; normalised text with the original tocorpus.pl [12] script that does standard normalisation; normalised with modified tocorpus.pl to remove one letter annotation tags; normalisation with modified tocorpus.pl to also remove more than one letter annotation tags; conversion of normalised text to uppercase; conversion to mixed lower and uppercase normalised sentence streams; and normalised text with added sentence boundary markers <s> and </s>. The versions were numbered 1, 2, 3.1, 3.2, 4.1, 4.2, and 5 respectively. The aim from this experiment was to arrive at the normalisation prescription that can be prescribed for LM development of the text for the chosen languages.

To determine the gain, if any, that may be derived from borrowing data from closely related languages, Experiment 2 combined or polled text data from different sources. For both languages, one combination is from both Lwazi and NCHLT texts, and the other is from texts of languages in the same cluster (Nguni or Sotho). Thus, the second combination will have further sub-combinations of Lwazi, NCHLT, and Lwazi+NCHLT.

In investigating high-order ngrams for the text of the two languages, the third experiment developed language models up to n-gram order 20. For each language, the experiment considered the different sized text in Lwazi, NCHLT, Lwazi+NCHLT, and cluster grouped text.

In analyzing the results from the three experiments, we observe that there is a need to approach language modelling of the two writing systems, conjunctive and disjunctive, differently. In particular, Experiment 1 (through Table 1) showed that normalized text with removed annotation tags develops better models.

From Experiment 2, by combining the Lwazi and NCHLT corpora, the recognition domain was widened and the size of the text corpus increased considerably. Table 2 indicates that, when compared to Experiment 1 NCHLT language models, there was an increase of about 3.9 in perplexity (PPL) for Pedi, and a significant drop of 10.3 for Ndebele.

High-order n-grams above the standard trigrams appear, through Figure 3, to be suitable for modelling both the under- resourced Pedi and Ndebele languages. However, the high order appears to be bounded to n = 6, after which the performance remains stable or deteriorates a bit throughout. For Pedi, low PPL ngram language models were pentagrams and hexagrams. Trigrams and quadrigrams were observed to better model the Ndebele text.

Looking forward, it will be interesting to observe and learn from the results of the other planned experiments, and the performance of the LMs in an ASR system.

Investigating the nature of public discourse in the #feesmustfall campaign

Sewchurran, Anusharani; Mdladla, Thulebona
University of KwaZulu-Natal

Abstract

There has been much in the traditional and digital media on the #feesmustfall movement. However, little appears in the academic space perhaps due to its currency. The initial focus was on the #Rhodesmustfall movement (Bosch 2016), with Booysen (2016) reviewing the #feesmustfall movement. Pillay (2016) also focussed on the movement from a psychological perspective. This paper will attempt to review the #feesmustfall campaign from a Habermasian perspective investigating the nature of public discourse emerging in the twitter feeds.

Widely acknowledged as one of the most prolific and influential theorists of our age, Habermas introduced and continually refashioned the idea of the public sphere (Habermas 1991) through which he contemplated the necessary conditions for 'mixed companies' or the populace to engage in rational critical debate which could then translate into political action. The theory of communicative action offers the ideal conditions of speech where such a public sphere could exist and out of which an authoritative political action could occur. Here reflexive agency and procedural rationality feature as necessary conditions towards a public sphere. Habermas's ideas of the public sphere were both critiqued and appropriated by a range theorists which led to the emergence of multiple competing publics existing as opposed to a cohesive singular metaphorical sphere (Calhoun 1992, Pusey 1987, Goode, 2005, McKee 2005, Edgar 2006 and Gripsrud, Moe, Molander and Murdock, 2010). In the face of competing publics Foucault's theories of power and discourse (Foucault 1997, 1998 and 2001) become significant especially within the context of digital media. While the media is not necessarily a public sphere, it does afford communication between and amongst 'mixed companies' thus creating the structural reality for public spheres to exist. Foucault's account of discursive fields is helpful in understanding the way in which competing publics jostle for the legitimacy to speak and for their truth to be valid, and validated.

The paper takes a qualitative approach using grounded theory. In this case the artefacts studied are the tweeter feeds related to #feesmustfall. Twitter analytics are guardedly employed. The study uses open coding with the constant comparative method to critically investigate the nature of the #feesmust fall discourses as they compete for validation and political action to occur.

An overview of available speech technology resources developed by the Human Language Technology Research Group of the CSIR Meraka Institute.

Calteaux, Karen; Titmus, Nina, De Wet, Febe
Council for Scientific and Industrial Research

Abstract

Human language technology development is a data intensive endeavour. Since 2003, the Human Language Technology Research Group (HLTRG) at the CSIR, Meraka Institute has been steadily contributing to the development of language resources required for speech technology development. This work has had a significant impact in the country due to the fact that the official languages, apart from English, are resource-scarce languages. This means that there are limited electronic text and/or audio corpora available for these languages, not to speak of parallel corpora. Nevertheless, through projects such as the National Centre for Human Language Technology (NCHLT): Speech and Text projects, funded by the Department of Arts and Culture (DAC), and initiatives such as the Resource Management Agency, good progress has been made in developing the resources required for language technology development in our languages. This paper provides an overview of the HLTRG's resource development work to date.

The paper begins by giving an overview of speech technology and contextualising the need for language resources for this type of work. While companies such as Google have developed systems such as Google voice search to assist with the collection of the resources required for language technology development, such systems do not yet exist in South Africa. A very brief mention of the HLT audit undertaken for the National HLT Network (NHN), will be provided to highlight the availability of resources at the time of the audit.

The paper will then briefly describe the speech technology work done by the HLTRG, to contextualise the resource development efforts undertaken to date. The HLTRG studies the way in which speech- and language-related technologies can be created and applied to benefit the people of southern Africa. Over the past 10 years, the HLTRG has been developing speech resources and technologies (ASR and TTS technologies) for South Africa's resource-scarce environment. The HLTRG at the CSIR Meraka Institute has been working on:

1. **Text-to-speech (TTS) technology development** for the 11 official languages of South Africa, where current research is focusing on speech prosody, code-switching and proper name pronunciation in TTS, and maturing the TTS technology into operational systems.

2. **Automatic speech recognition technology development** for the 11 South African official languages where current research is on improving in-domain ASR performance and developing human capital in ASR research and development.

3. **Resource development** for the 11 official South African languages

a. The efforts began in 2006-2009 with annotated speech corpora for 11 languages, with close to 6 hours of telephone speech per language in the ASR corpora and almost 1 hour of speech data per language in the TTS corpora.

b. By the end of 2012, around 60 hours of broadband speech had been collected for each of the 11 official languages in the National Centre

for Human Language Technologies Speech Resource Development project.

c. From 2013 to date, the focus has been on extending the NCHLT I speech corpora by automatically enhancing the annotation level of existing multilingual data sets.

The paper will provide details on the types of corpora available, descriptions of these, their terms of use, and how to access them, in as much detail as is possible within the time constraint. The paper will conclude with observations on possible future speech resource development work to be undertaken, based on needs that have been identified by the HLTRG in conjunction with industry partners and government departments. Some challenges in this regard may also be touched on briefly.

Combining task-based teaching and computer assisted language learning (CALL) in a computer program for international students studying Afrikaans at Stellenbosch University

Beukes, Vernita
Stellenbosch University

Abstract

This paper focuses on two learning areas namely task-based learning and computer assisted language learning (CALL) and indicates how these two learning areas can be combined in a computer program for international students studying Afrikaans at Stellenbosch University. The question of how these two learning areas could be used in a South African context with Afrikaans as the instructional language, was used to underpin the research.

In my presentation, I will firstly provide some background on both the Beginner Afrikaans level 1 course as well as the students who are enrolled for the course. The above mentioned course falls under the International Program at Stellenbosch University (IPSU) which is coordinated by die International Office. The Language Centre at Stellenbosch University is the service provider for all language acquisition courses.

The presentation will further discuss the cognitive theory of multi-media learning developed by Richard Mayer (2001, 2005, 2009). This theory investigates how a multi-media presentation containing text, video and pictures enters the sensory memory, gets processes by the working memory and how it gets integrated with prior knowledge so that it can be stored in the long-term memory. The theory further emphasises the importance of selecting and organising sounds and images so that it can become a verbal model and pictorial model. Richard Mayer (2001) also developed twelve instructional principles as part of his evidence-seeking efforts for the science of instruction which can be used to design any multi-media presentation. The main aim of these principles is to reduce extraneous processing, to manage essential processing and to foster general processing. I have chosen four of the abovementioned instructional principles namely the coherence principle, the redundancy principle, the spatial contiguity principle and the personalization principle which will be used in the computer program.

A needs analysis asking students to list the subjects that they want to be able to talk about was also done with the students who attended the Beginner Afrikaans level I course. Out of this abovementioned needs analysis 8 subjects was identified. The subjects (in order of merit) was:

1. Places to visit in South Africa

2. Direction

3. Housing

4. Hobbies

5. Ordering food

6. Weekend plans

7. Buying food in a supermarket

8. Movies / Favourite Actors / Favourite Actresses

The four most popular subjects namely places to visit in South Africa, direction, housing, and hobbies was taken and combined with four of the design principles that Mayer (2001, 2005,

2009) suggested. Focus on form activities as stipulated by Willis (1996, 2007) was also added to the four design principles so that the connection with task based

teaching could be achieved. The computer program that was developed out of the abovementioned will be discussed in detail.

The last part of the presentation will focus on the analytical framework that I intend to use to evaluate the computer program that was developed at the end of 2016. I have decided to apply certain aspects of expansive learning because it focuses primarily on change and enhancement of any program. Expansive learning was developed by Engeström (2001).

In my presentation I will shortly explain the seven steps within expansive learning: In steps 1 and 2 of expansive learning an ethnographic analysis of the current situation will take place in which the students will question their present activity by jointly analyzing problematic situations in the computer program. According to steps 3 and 4 of expansive learning the transforming and examinig of the computer program should then take place in order to find a new and improved version of the program. The new program should then according to step 5 of expansive learning be implemented and the new version should be tested. In steps 6 and 7 of expansive learning reflection on the new program and consolidation of the new program should take place.

In 2017 I intend to implement the analytic framework by focusing on steps 1 and 2 in the first semester. Students that attend the Beginner Afrikaans level 1 course are only here for one semester and therefore the students that attend the abovementioned course in the first semester (February 2017 to May 2017) will fill in a questionnaire and be part of interviews to determine the problematic situations within the computer program. The program will then be refined between the first and second semester. The students that attend the course in the second semester (July 2017 to October 2017) will then test and reflect on the new version of the program.

Using an online marking tool in developing and assessing academic literacy

Meihuizen, Elizabeth Maria
North-West University

Abstract

In this paper a corpus of writing by first year university students is used to determine the value and best mode of employment of an online marking tool in developing and assessing academic literacy abilities.

The online marking tool Schrijfhulp Nederlands was developed by researchers in the Instituut voor Levende Talen of the Katholieke Universiteit Leuven in cooperation with other institutes of higher learning. The user can upload text and request feedback and is then provided with assistance in improving the writing. The assistance comprises two main categories, namely text editing and text enrichment. Text editing is provided with regard to coherence and cohesion, style, and spelling. Under the rubric text enrichment the writer is provided with support such as looking up the meanings of words, choosing from a provided list of alternative terms suitable for use in an academic environment, and checking for the use of words regarding context and combination with other words. The developers of Schrijfhulp Nederlands stress the fact that, although support is provided by their marking tool, the user still maintains autonomy and remains responsible for improving the writing. This also implies that the support provided focuses on the process of writing rather than on once-off text production.

The developers of Schrijfhulp Nederlands are in the process of testing and refining their marking tool, and for this purpose have made an English version available to some South African universities for evaluation. This paper reports on work done in this regard in the Centre for Academic and Professional Language Practice at the Potchefstroom Campus of North-West University. The focus of this paper is on essay type student writing submitted as part of the course in Academic literacy. Feedback provided by the lecturer on an assignment that forms part of a more extended process of developing academic writing skills is compared to that provided by the online marking tool. The aim is to determine the potential contribution of the online marking tool to existing modes of assessment and instruction and to suggest further development of features or refinement of existing ones.

Adequate and appropriate? Matters to consider for the validation of an online writing tool.

Van Dyk, Tobie1; de Wachter, Lieve2; Verlinde, Serge2
North-West University1, KULeuven2

Abstract

It is generally acknowledged that secondary education systems across the world do not necessarily prepare students adequately and appropriately for the demands of higher education. There is agreement among authors that a lack of proficiency in the language(s) of teaching and learning, particularly in the productive mode, has a detrimental effect on study success. First-year students are predominantly at risk of failure and universities attempt to address this in many ways, among others by offering different kinds of interventions to help students better integrate and acculturate to the academic environment. One such intervention is the establishment of writing centres that students can visit free of charge for a one-on-one consultation with a trained writing consultant. However, as the demand on writing centres seems to increase and the availability of resources is limited due to cost or time factors, other initiatives need to be put into place to guide students textually before visiting the writing centre. An initiative taken by the Instituut voor levende talen, KU Leuven was the development of an online writing tool available to all students and staff associated with this institution. The tool is available in both English and Dutch. Users can upload their text to the online platform and get feedback on (i) structure and coherence, (ii) style, and (iii) spelling and grammar. Each of these fields

consist a number of components that check aspects of students' written work and accordingly provide information and guidance on how to deal with the matter at hand. The purpose of this paper is to investigate the validity of the aforementioned tool at a South African university. Matters considered include primarily the efficacy of the tool and its face and content validity. The former will consider possibilities for refinement and implementation. The latter will compare lecturer opinion with that of the tool in terms of the three areas of feedback mentioned above. The paper concludes with a brief overview of challenges and opportunities as to inform decision making on further investment.

Digital humanities practice: Site-specific digital literature

Greyling, Franci
North-West University

Abstract

The field of electronic literature (e-literature, digital literature) which emerged at an "interdisciplinary nexus of literature, creative writing, communications, computer science, art and design" (Rettberg 2016), is increasingly associated with digital humanities. Scott Rettberg (2016) describes the creative production of electronic literature as a digital humanities practice: "not an application of digital tools to a traditional form of humanities research, but rather experiments in the creation of new forms native to the digital environment". The rapidly changing digital environment offers writers and artists abundant opportunities for experimentation, as is evident in the diverse forms and threads of practice, such as hypertext fiction and poetry, kinetic poetry, literary apps, interactive fiction, collaborative writing projects, computer generated literature, literature in the form of social media, and site-specific digital literature. In this dynamic environment, there is also the need for approaches and critical language to understand, describe and theorise the developing narrative forms and practices.

In order to investigate these new creative and publication possibilities, an interdisciplinary practice-based research space (Byderhand) was established for the experimentation with and understanding of digital literature – especially site-specific literature. Site-specific digital literature (locative literature, place-bound literature, ambient literature) is a form of electronic literature in which multimodal texts are made available at a specific site by means of digital (especially mobile) devices. The concrete place is inherent to the creation as well as the reading or experiencing of these texts. Some international examples of site specific digital literature projects are textopia (a project by Anders Løvlie in Oslo, Norway), StoryTrek (a project developed by Brian Greenspan in Ottowa, Canada), and StreetReads and the Story City app (produced by Emily Craven in Adelaide and Brisbane, Australia). Byderhand 2015 was presented as part of the literature programme at the Clover Aardklop National Arts Festival in Potchefstroom, South Africa and was marketed as an interactive reading festival. The project consisted of four sub-projects involving different locations, genres and target readers, namely: poetry and children's poetry in the Botanical Garden, short stories and children's stories in a coffee shop, a story for teenagers on a local school ground, and comic strips in taxis. Readers accessed the multimodal texts by means of QR-codes scanned with their cell phones. In the poetry project (Byderhand Tuinverse) fifteen poets wrote poems for the North-West University Botanical Gardens. The poems are complemented by recordings of readings by the poets, typographic animations and musical arrangements. The user interface enables the reader to select and experience a particular rendition of a poem. The artist Strijdom van der Merwe made 15 totems to be displayed with the poems, thus combining digital art and tangible art. In 2016 translations of the poems in English and Setswana were added to the selection. For more information on the project see www.byderhand.net.

Byderhand 2015 could be considered as an experimental publication system and involved the aspects of production, marketing and

distribution, mediation and reception of texts. Rettberg (2011) argues that since a publishing industry for electronic literature doesn't yet exist, the collaborative effort in the creation, publication and distribution of electronic literature is more clearly evident than in the traditional publishing industry where the collaboration has become invisible to a large extent. In experimental literature the roles of designers, artists and editors are usually more clearly identified and acknowledged. The post-project reflection of Byderhand 2015 confirmed that this liminal zone is indeed a good place to experiment with and to describe processes, collaboration and new cultural forms.

The question examined in this paper is how the interaction and interplay in this experimental publication system for site- specific digital literature can be understood and represented, so that it can be of value to: 1) the conceptualisation and presentation of site-specific digital literature projects; 2) the understanding, description and analysis of multimodal site- specific digital literature. A multimodal ensemble for site-specific literature has been compiled, which relied on Miller's (2011) key elements of digital media, Kress and Van Leeuwen's (2001) multimodal theory of communication, and Page's (2011) multimodal ensemble for narrative analysis. Three dimensions of the composition and orchestration are distinguished in the proposed multimodal ensemble for site-specific digital literature, namely: 1) the components of site-specific digital literature; 2) communicative practices and technical processes; and 3) immersive experiences.

Digital and crucial exclusivities

Sewchurran, Anusharani
University of KwaZulu-Natal

Abstract

It is widely acknowledged that critical thinking is one of the most important skills for human sciences students to develop. This is both in terms of general life skills and for survival in the modern marketplace (Bozalek and Watters 2014, Dongwe 2013, Dunne 2013, Jacobs 2007, Paxton

and Frith 2014, Rossouw 2003, and Shalem and Pendlebury 2010). The imperative for critical thinking is also strong in national (the center for higher education) and institutional policy discourses (UKZN policy on teaching and learning). The amalgam of massified tertiary education, issues around access, cost factors and pressure to research most, have led academics to consider digital solutions as a means of achieving teaching goals. Institutions often compel academics to use the digital space more (e.g. phasing in of moodle at undergraduate years – UKZN).

This paper interrogates digital affordances for the tertiary teaching context in relation to developing critical thinking. The paper problematizes superficial digital solutions as means of engagement. It reflects on some of the digital imperatives upon a university teaching context and explores alternate ways towards authentic learning within and alongside a digital environment. The study departs from traditional notions of critical thinking to critical being (a modern variety of Aristotle's phronesis) which is encompassed in the notion of 'criticality' articulated by Dunne (2014). The paper emerges from Whiteheads 'Living theory research' (Whitehead and Huxtable 2013) which offers a fundamental question as a reflective form of self-study, 'how can I improve what I am doing?' Whitehead asserts this question as a legitimate starting point to "an ethically driven form of research where the educator recognizes and takes responsibility for the contribution they make to the quality of the educational relationship, space and opportunities experienced by learners" (Whitehead 2013, 2).

Many tertiary education scholars have reflected on the changing higher education landscape in South Africa (Cloete, Fehnel, Maassen, Moja, Perold and Gibbon 2002, Singh 2015, Rossouw 2003, Simelane, Bignaut & Van Ryneveld 2007, and Moodley 2015). Tertiary level teaching is in a manner of speaking, under siege by a range of pressures from globalization, competition from private entities, rapidly changing work environments requiring newly and differently skilled graduates (Cloete, Fehnel, Maassen, Moja,

Perold and Gibbon 2002, Singh 2015 and Moodley 2015) and the intensified appropriation of attention by the social media (Davenport and Beck 2002). While some scholars have focused on structural transformation, others have focused on the changing technological landscape (Blewett 2015, Bozalek & V Ng'ambi 2013, Bozalek, & Watters 2014, Cahalan 2013). It is clear that the new globalized marketplace is complex, diverse and mutable. It has resulted in the transformation of many traditional workspaces. In the modern workplace jobs are conflated as a result of dematerialization and old technological knowledge is rapidly rendered obsolete with traditional jobs being eroded from the work-scape (Glenn 2010, 254-255). This climate coupled with the economic downturn has raised interesting questions around tertiary education particularly for the Human Sciences (within which Media and Cultural Studies resides). Media and Cultural Studies as a discipline is perhaps the most affected by a rapidly changing technological landscape. Alongside this the work of media graduates is also rapidly changing.

In preparing graduates for this changing workplace one needs to consider both critical thinking and engaged citizenry as necessary conditions to survival of the individual and the larger democratic order however fragile. Bozalek and Watters (2014, 1070) state that, "graduate attributes…should allow graduates to learn 'for an unknown future'. Dunne argues for criticality as a 'raison d' etre of higher education (2014, 86). He states that criticality must be differentiated from critical thinking as it encompasses, "critical thinking, analytical reasoning, critical self-reflection and critical action" (2014, 87). This is a fundamental aspect to preparing graduates for new workplaces and employers lament the lack of critical thinking, writing and problem-solving skills (Dunne 2014, 87). Dunne while referring to the former president of Harvard University argues further that there are 'yawning chasms between the rhetoric of educational policy and the verifiable actuality of student learning' (ibid). This means that while critical thinking is highlighted at the policy level, it is not necessarily realized in teaching. Bozelak and Watters (2014, 1070) indicate that while this discourse around

graduate attributes and criticality is important, "one of the problems of looking solely at curriculum alignment to achieve graduate attributes is that many of the attributes are process driven." They offer an alternative approach to achieving graduate attributes which is a shifted focus to 'authentic learning' and 'authentic learning environments (ibid).

This paper addresses the complex translation of policy imperatives (with a specific emphasis on the "critical") into meaningful (digital) practice. The paper engages with the processes of embedding criticality as articulated by Dunne (2014) and graduate attributes and authentic learning (Bozelak and Watters 2014) in a larger massified, digital context.

The rioting writer: understanding fees must fall as design within Transhumanist South African Academia

Maritz, Lindi Helene
Greenside Design Center, College of Design

Abstract

The dream of transcending human limitations is not a new desire neither is the call for innovation within academia and scholarly activities. With ever expanding technologies, digital approaches and design(ed) methodologies the fruition of these desires has never been closer. Yet, we cannot proceed blindly towards this promise of progress.

It seems impossible to imagine that any factuality of academia in South Africa has remained untouched by the effects of the Fees Must Fall (FMF) demonstrations. The student led protest began as a reaction towards the increase in fees by South African Universities in 2015. Even though there were no fee increases in 2016 the protest started again when the Minister of Higher Education announced a maximum of 8% increase for 2017. South African education is fused with the actualities of power struggles and FMF eerily echoes the events of the June 16 (1976) Soweto uprising. FMF once more

highlighted the agency and insight of South African youth. However the project of FMF seems to have progressed beyond achievable education for all towards an uncertain outcome.

The terms trans-humanism and post-humanism are regularly used interchangeably, however the ethos and outcomes are vastly different. Transhumanism can be defined as a cultural movement that affirms the desirable possibility of improving the human condition through applied reason and explicate technologies. This upgrade does not only concern the human body but also mental capacities (Bostorm, 1999). This advancement is there for a project of overcoming the body as a defective and the enhancement of the physiological as desirable, or in other words a call for reason. Post- humanism refers to the conception of human transcended from traditional boundaries through technological augmentation. However, post-humanism questions whether we are already post-human. The notion that we became (Hayles 1999) post-human is due to the impossible imaging of human without technology or clear demarcation of human and species or nature. The commonalities between post-humanism and trans-humanism pertain to three seminal concepts; human (mind), technology and nature (body). Post-humanism sees these concepts as points in a system. Transhumanism distinguishes the body as nature and technology as a means to enhance the mind. Post-humanism proposes a flatting of binaries within a triad, but in no way does this flattening propose a harmonious co-existence.

It is important to note that trans- and post-humanism's preoccupation with technology does not always acknowledge or address the 'not yet technological liberal societies' and that qualities identified for improvement are under written by imperialist and western ideals. Furthermore, trans- and post-humanism raises issues of body politics as well as Cartesian subject hood. As such, this research proposes that the 'second wave' (2017) of FMF is a violent regurgitation of Eurocentric, trans-humanists project of enlightenment and progress.

The research positions the activities of (digital) design not as instrumental or as neutral mediation, much like technology, between client and audience. The practice of design is rather an act of critical and reflective writing. Post-human other Hayles (1999) states that the practices of reading and writing within the digital media environment and digital technology

have created new conditions for concepts of identity and subjectivity It can be said that both writing and designing requires research, creative and critical thinking on the subject and execution of several drafts that are refined as the conceptual focus evolves. Thus, by interpreting subjects a designer must consider multiple view points and the role of representation considering "life-centred" ideas (Mc Carthy 2002/11). By connecting pictorial, verbal as embodied and digital languages through mass communication channels, such as computer mediated communication technology, requires the designer to critically examine the social and cultural role of themselves within a larger systems. Therefor digital design considers specific productions and performances and matches them to different methods in a way that pushes publics to re-examine alternatives. It is through this lens that the research is unpacked.

At the 2016 triennial conference of UNISA's school of Arts; Professor Pumla Gqola stated in her key note address that we can no longer distinguish between the writer and the rioter as we re-map, re-think and re-imagine Africa. She also states that in an African locale the exclusive perception that the act of writing as safe and the act of rioting as dangerous is a fiction. In doing so the conflict between the students and the education system is re-structure and agency is dispersed. Students are reassembled as educators and educators as students. This also corresponds to the awareness that the educator does not stand outside the praxis of design, but is a designer themselves. The divide between writer and rioter, design and user, student and educator becomes ambiguous.

The research aims to explore the precepts of trans-humanism and its relation to the post-humanism to understand the social, economic

and political implications of FMF. The research also questions whether the relation between the online hashtag activism of #FeesMustFall and the offline protest of FMF, specifically in 2016, is a manifestation of the digital divide within developing and third world countries? Lastly the research aims to analyses literature against the complex and possible contradicting priorities of the subject matters within this research. The intention of the research is not to resolve issues within the FMF movement in a manner that echoes Eurocentric trans-human liberal enlightenment. Neither is it to resign with ambiguity, which is often related to the conclusions of post-humanism dialog. But rather to critically engage with the shifting paradigms with in digital design, education and humanities in South Africa, Africa and the global village at large.

The Five Hundred Year Archive: Building a Digital Archive for the Southern African Past before Colonialism

Hamilton, Carolyn; McNulty, Grant
University of Cape Town

Abstract

The southern African past before the advent of European colonialism remains one of the most under-researched aspects of the history of the region. There are several reasons for this, of which two stand out. Firstly, while some of the relevant resources are text-based, many exist in other forms, the archival potential of which is not readily apprehended. The second reason is the way in which the materials pertinent to the remote past have been shaped by disciplinary conventions (such as museology, archaeology, anthropology etc.) and colonial and apartheid knowledge practices, to be treated as timeless, traditional and tribal materials.

A key move in the Five Hundred Year Archive (FHYA) project is the creation of a digital exemplar (in the form of a web platform), which is capable of bringing together digitally, textual, visual and sonic materials pertinent to these periods. The exemplar aims to be a conceptually innovative intervention geared to engaging, in a critical manner, inherited forms of knowledge organisation. It is being constructed to work across multiple institutions and disciplines, to include multimedia (digital text, images and audio) and to provide context, by taking into account the provenance of the various materials, as well as their multiple histories, and how they have changed over time.

This paper presents some of the challenges involved in translating the conceptual thinking behind the FHYA into a practical project. It proposes a standardised framework for the FHYA Exemplar that can accommodate the variety of materials emanating from different partner institutions and disciplines.

Contextualization Tools for Digital Archives

Escobar Varela, Miguel
National University of Singapore

Abstract

This presentation argues that providing contextual information is a key feature of digital archives. Three case studies are used to illustrate conceptual and technical challenges arising from this feature.

In the digital humanities, digital archives often blend features of archiving and editing (Price 2009). The primary mission of archives is not to present a consistent impartial group of records, but to offer contextual information for a large selection of records. Perhaps the archiving principle of respect des fonds [respect for the structure of the records] should be reimagined as context des fonds [context of the records]. This paper considers the implications of this new archiving maxim by analyzing three digital archives of Asian theatre performances: the Asian Shakespeare Intercultural Archive (A|S|I|A), The Contemporary Wayang Archive (CWA) and Theatre Makers Asia (TMA). Although all three archives were developed jointly at the National University of Singapore, they each respond to different needs in archival practice

and include different contextualization tools. A contextualization tool is defined here as a widget, embedded panel or interactive display of a website that provides context for the users of an online archive.

The Contemporary Wayang Archive (CWA) is a collection of re-elaborations of Java's oldest and most important performance tradition: wayang kulit [leather puppets]. It includes performance data, translations and notes on how the performances were received in their original context. Besides conventional metadata (author, place and time of recording), each performance record includes a short summary and interpretive tags that specify the type of music, performance space, puppets used and languages spoken in each performance. This metadata forms part of a digital ecosystem of contextualization, that can be accessed by the users in several ways:

1) A list of all the recordings, where selected metadata tags are displayed. A user can hover the mouse over the tags for a description.

2) Individual recording pages, where all metadata items and the summary are displayed. The video can be played and a side panel includes notes on the translation of the performances, as well as a list of honorifics

(Fig. 1). The notes include word play, cultural and historical allusions, and ambiguous expressions are explained in the notes. A number in brackets in the subtitles refers to the note number that can be found in the note box next to the video player. The translations aim for simplicity in order to allow users to easily watch the performance videos. However, the honorifics in the original languages were retained in the subtitles. In Javanese and Indonesian, speakers address their interlocutors with over 40 different honorifics which denote differences in their relative status and level of intimacy.

3) A (forthcoming) SPARQL endpoint where the semantic data can be queried directly or linked into federated data collections.

The archive provides interaction paths for different kinds of users. People with little expertise in Javanese theatre can read introductory sections, use the tag navigator to find performances and watch the videos while reading the contextual notes and honorifics. Advanced users can use the concordance search to find specific passages in the transcripts, or watch the performances in full screen mode, with transcripts or with no subtitles. These contextual tools can be turned on and off according to the users' needs.

The objective of A|S|I|A is to provide "the basis for an accurate, contextualized and comparative experience of Asian Shakespeare performances." This approach is implemented in the main contextualization tools of the archive: interactive concept maps. These maps reflect culturally-specific choices in the theatrical forms, points of reference, staging strategies and approaches to adapting Shakespeare in over sixty performances. The high degree of nuance in these data points can be explored through the interactive map in relation to specific performances. However, the aggregate of this interpretive data can also be visualized through comparative pie charts that show the distribution of categories across the records in the archive (Fig.2).

Another contextualization tool is the presence of multiple languages: Chinese, Japanese, Korean and English. The A|S|I|A website itself represents an exploration of intercultural practices for the digital archiving of performance since its four parallel language environments are fully equivalent and users can switch between them at any time. The subtitles and contextual data are consistently translated across these languages. In other words, the kind of context provided by the archive is not generic for all website visitors, but tailored to the language preferences of individual users. Lastly, users can add their own data through another contextualization tool: a time-based annotation panel, where users can comment on specific moments of a performance. These comments can be made public or restricted to private groups. But in either case, they provide the users with a way to augment the archive with their own set of contextual notes.

TMA, in contrast, aims to experiment with digital archives as platforms for collaborative research and creative responses. The collection of the archive focuses on the history of contemporary theatre in East and Southeast Asia. Rather than organize the archive material according to invariant structural principles, the records can be organized into exhibitions. Some of these will be curated by guest researchers, but individual users can create their own exhibitions. The exhibition is another kind of contextual tool, where context is dynamically recreated as the result of considering the performances in dialogue with one another.

As seen in this overview, contextualization poses technical and conceptual challenges for digital archives. Conceptually, it forces the makers of archives to interrogate the ontology of their creations: are they collecting comprehensive storehouses of records or scholarly introductions to little-known themed collections? The answers to this question can lead to a wide array of technical implementations of contextual tools. Even within a narrow field, such as Asian performances, a single solution will be impossible. The future of archives as useful resources lies in a plurality of approaches.

Works Cited

Kenneth M. Price (2009). Edition, Project, Database, Archive, Thematic Research Collection: What's in a Name? Digital Humanities Quarterly Volume 3, Number 3.

Digitising Herman Charles Bosman – An initiative based on special collections projects by the North-West University's Library Services (Potchefstroom Campus).

Van Deventer, Kirchner; Coetzee, Anneke
North-West University

Abstract

Digitisation of hardcopy documents, books, correspondence and other data is increasingly being used by institutions such as libraries to create and deliver digital representations, in order to define new avenues of understanding in the humanities (Terras, 2012:48). Consequently, institutions of higher learning are increasingly committing resources to digitisation projects, since they enhance learning and research and could ultimately have a profound impact on the quality of output these institutions provide (Terras, 2012:49). Moreover, digitisation of any collection allows for closer collaboration between institutions that may be thousands of kilometres apart (Carter, 2013:97). Consequently, any endeavour that an institution of higher learning might undertake to digitise a particular collection, would be worth the expense, time and labour it might entail.

In line with the above, the North-West University's (NWU) Library Services (Potchefstroom Campus) has been involved in several special digitisation projects of its more valuable collections. These projects include, but are not limited to, the Bibliography of the Afrikaans Language (DBAT); ISAM (Information Centre for Southern African Music); the Pierneef collection which consists of information about and digital copies of the works by landscape artist, Jacobus Hendrik (Henk) Pierneef; the Cloete manuscripts which is a restricted access database containing manuscripts that were donated by the family of the late T. T. Cloete; lastly, the Ossewabrandwag collection which contains digitised archival material related to this mid-20th-century movement. In addition to the above mentioned digitisation projects, the NWU will be hosting a multi- partner South African Centre for Digital Language Resources, that is tasked with creating and managing digital resources and software that can be applied to research related to all of South Africa's official languages (Department of Science and Technology, 2016:36). This initiative, which will be the first of its kind for Africa will provide ground-breaking research and development based on Digital Humanities. Consequently, given its past involvement with digitisation projects, it is envisaged that not just Library Services (Potchefstroom Campus), but all the NWU Library Services will be involved in future Digital Humanities projects.

Given the above, NWU Library Services will become not just willing partners in Digital Humanities (DH) initiatives but, more importantly, should be active participants and initiators of DH projects that will enhance and promote research and development in this field, whilst collaborating with the Faculty of Arts and Humanities in prioritising projects.

Therefore, NWU's Library Services has endeavoured to initiate a digitisation project of the collected works and all other available materials of Herman Charles Bosman. Bosman was a South African author who was born in 1905. He began his career as a teacher in the Groot Marico district, which became the setting for many of his stories. He also worked as a journalist and literary editor, while also publishing his own work (van Bart, 2016). Needless to say, the breadth of his own work and other resultant documents, articles, court cases, etc. has provided a profound amount of literary and linguistic research. The NWU Potchefstroom's School of Languages recognises this; and their interest stems from the fact that Bosman spent much of his youth in Potchefstroom, which is situated in the North-West Province and also is the setting for many of his stories (van Zyl, 2016). Moreover, Bosman's particular literary style incorporated both Afrikaans and English into much of his work and, hence, the linguistic and literary interest in him (van Zyl, 2016). A comprehensive digital repository of Bosman will, consequently, drive the DH agenda of the NWU.

With any new project involving the digitisation of hardcopy materials, several factors need to be considered. These include feasibility, cost, labour, access to and preservation of materials, copyright matters, available technological software and hardware infrastructure, and many more. In order to address these factors in the proposed presentation, an overview of the existing special digital collections of the NWU's Campus Libraries will be provided by the presenting librarians. The

overview will assess how these projects addressed the factors above. In addition, possible hindrances or benefits will be identified that can

be utilised for future digitisation projects, such as the proposed Bosman digital repository.

The presentation will therefore continue by outlining a road map for the proposed Bosman digitisation project, beginning with an overview of what has been accomplished during the period September 2016 to January 2017. The road map will continue by addressing concerns regarding the factors mentioned above and how each factor will be approached going forward.

In summary, the presentation will conclude with considerations regarding the viability of the project and, lastly, how it can contribute to increased cooperation between academic libraries and Digital Humanities initiatives.

Technological Adoption and the Role of Cloud Computing in Digitization, Knowledge Creation and Digital Curation in Africa.

Matlhako,M. Mamadi; Ajibade, Patrick
University of Fort Hare

Abstract

In the 21st century, there has been a pervasive occurrence of Information Technology (IT), and this has impacted the creation of digital collections and the ways knowledge is created, use and share. This technology has undoubtedly promoted access to relevant archival materials. However, it has presented pockets of security challenges with the adoption of computer technologies in producing digital content in an electronic and networked environment. The aim of this paper is to engage and encourage debates on the opportunities, and challenges of an optimised framework for digitisation. It is noteworthy to consider the implication of latest trends in information technology such as cloud computing vis-à-vis digital curation and digital preservations. The paper examines the levels of current technological infrastructure in Africa, the technical and organisational requirements to implement knowledge archiving in the cloud and highlight the current challenges in cloud computing. This may be useful in digital curation

and digitisation. Analysis of the dynamic interaction and computer architectures' implication on digitisation will be discussed. The paper address issues surrounding its implementations, suitability and the different stack control. The presentation will highlight institution's concerns on service capability provided, cloud applications, and choice of platforms or infrastructures. These specification and service requirements are crucial for performance agility. Service and deployment models, technological infrastructures for born-digital knowledge preservation, as well as the integrity of the collections. Security, limitations, suggestion and recommendations are provided to enable institutions to develop reliable models that meet their organisational goals.

Digitising photographic collections in the Western Cape, South Africa – Key issues affecting image quality and digital quality management for preservation purposes.

Carstens, Antenie
National Library of South Africa

Abstract

This information for the oral presentation, formed part of a recent Master's study completed at the Cape Peninsula University of Technology.

This study deals with the pitfalls and standards associated with the digitisation of photographic artefacts in formal collections. The popularity of the digital medium caused a rapid increase in the demand for converting images into digital files. The need for equipment capable of executing the task successfully, the pressure on collection managers to display their collections to the world and the demand for knowledge needed by managers and operators created pressure to perform optimally and often in great haste.

As a result of the rush to create digital image files to be displayed and to be preserved, the decisions that are being made may be questionable. The best choice of file formats for longevity, setting and maintaining standards to

guarantee quality digital files and consultation with experts in the field of digitisation as well as attention to best practices are important aspects which must be considered.

In order to determine the state of affairs in countries with an advanced knowledge and experience in the field of digitisation, a comprehensive literature study was done. It was found that enough information exists to enable collection managers in South Africa to make well informed decisions to ensure a high quality of digital collection.

By means of questionnaires, a survey was undertaken amongst selected Western Cape image preservation institutions to determine the level of knowledge of the managers who are required to make informed decisions. The questionnaire was designed to give insight into choices being made regarding the technical quality, workflow and best practice aspects of digitisation. Comparing the outcome of the questionnaires with best practices and recommended standards in countries with an advanced level of experience it was found that not enough of this experience and knowledge is used by local collection managers although readily available. In some cases standards are disregarded completely.

The study also investigated by means of questionnaires the perception of the digital preservation of image files by fulltime photographic students and volunteer members of the Photographic Society of South Africa. It was found that uncertainty exist within both groups with regard to file longevity and access to files in five to ten year's time. Digitisation standards are set and maintained by the use of specially designed targets which enable digitising managers to maintain control over the quality of the digital content as well as monitoring of equipment performance. The use of these targets to set standards were investigated and found to be an accurate and easy method of maintaining control over the standard and quality of digital files.

Suppliers of digitising equipment very often market their equipment as being of a high quality

and being able to fulfil the required digitisation tasks. Testing selected digitising equipment by means of specially designed targets proved however that potential buyers of equipment in the high cost range should be very cautious about suppliers' claims without proof of performance. Using targets to verify performance should be a routine check before any purchase.