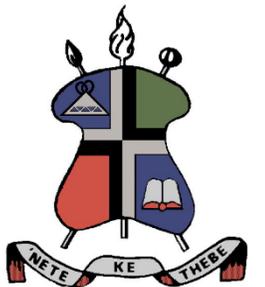


# African Languages in the Age of 4IR

Machine Learning and Automated Translation Systems for African  
Languages

**Lehlohonolo Mohasi**  
National University of Lesotho



# Opening Statement

We are in the fast-moving era, the 4IR, which basically means technology is at a stage where things have become smaller, faster, etc. Language, being a part of human nature, must not be left behind. Our languages, African languages in particular, must not lag behind. A lot of research has been done on European/American/Asian languages, and more must still be done for African languages. We must also be cognisant of minority languages which must not die. According to Emmanuel Gabriel, the founder of Germany-based OpenBinacle and creator of OBTranslate, about 52 native languages in Africa have undergone language death. In our undertaking of research and chasing technology therefore, no language must be forgotten. Language is the core of any nation's culture, and there is no nation without culture.



# Machine Learning

- Machine learning is an application of artificial intelligence (AI) that provides systems the ability to *automatically learn and improve from experience without being explicitly programmed*. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.
- The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to *look for patterns in data and make better decisions in the future* based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.



# Machine Learning Algorithms

- **Supervised:** Apply what has been learned in the past to new data using labelled examples to predict future events.
- **Unsupervised:** Used when the information used to train is neither classified nor labelled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabelled data.
- **Semi-supervised:** Use both labelled and unlabelled data for training – typically a small amount of labelled data and a large amount of unlabelled data. The systems that use this method are able to considerably improve learning accuracy.
- **Reinforcement:** A learning method that interacts with its environment by producing actions and discovers errors or rewards. Trial and error search and delayed reward are the most relevant characteristics of reinforcement learning.

Machine learning enables analysis of massive quantities of data > Requires additional time and resources to train it properly.

(<https://expertsystem.com/machine-learning-definition/>)



# Machine Learning Applications

- The heavily hyped, self-driving Google car
  - The essence of machine learning.
- Online recommendation offers such as those from Amazon and Netflix
  - Machine learning applications for everyday life.
- Knowing what customers are saying about you on Twitter
  - Machine learning combined with linguistic rule creation.
- Fraud detection
  - One of the more obvious, important uses in our world today.



# Machine Learning Importance

- Growing volumes,
- Varieties of available data,
- Cheaper and more powerful computational processing,
- Affordable data storage.

All the above mean it's possible to quickly and automatically produce models that can analyse bigger, more complex data and deliver faster, more accurate results – even on a very large scale.

What's required to create good machine learning systems?

- Data preparation capabilities,
- Algorithms – basic and advanced,
- Automation and iterative processes,
- Scalability,
- Ensemble modelling.



# Automated/Machine Translation

- Machine translation software automates the process of translating text from one language to another. The most basic machine translation software relies strictly on word-for-word substitution.
- Some technologies include rule-based or statistically-modelled translation for more accurate translations.
- Machine translation tools translate text without human editing.



# Machine Translation Category

- Translate text word for word, or incorporate rule or statistic-based functions;
- Automate translation into multiple languages;
- Instantly produce copy that can be edited by translators;
- Provide tools for editing translated text;
- Organize or allow for the management of translated text.



# Machine Translation Software: Top 7

- Google Translate
- Microsoft Translator
- Yandex
- Amazon Translate
- IBM Watson Language Translator
- Cloud Translation API
- Bing Translator

(<https://www.g2.com/categories/machine-translation>)



# Translation Tools Types

Translation tools can roughly be divided into three types:

- Fully Automated Machine Translation (FAMT)
- Human Aided Machine Translation (HAMT)
- Machine Aided Human Translation (MAHT)

The criteria above are centred on two aspects:

- Translation tool's degree of automation, i.e. the degree to which the machine/software independently conducts the actual translation.
- Who controls the translation process, i.e. who – the software or the translator – controls which decisions are made in the process.



# Fully Automated Machine Translation

- Fully automated machine translation (or simply Machine Translation or MT) is understood as software capable of conducting a flawless translation between natural languages independent of human interference or help. The human role has been reduced to simply loading the desired text into the computer.
- The different types of FAMT-programs
  - The Direct MT-Model
  - The Transfer Model
  - The Interlingua Model



# Human Aided Machine Translation

- Software developed for the machine to translate what it can, in the way it can. The human role can be compared to that of a consultant or an editor, i.e. that the translator corrects or modifies what, in the machine's translation suggestions, is unacceptable to him or her.
- SYSTRAN is an example of a HAMT-program.
- Human translator can, in principle, take on his/her consultant- or editor-role *before, during* or *after* the machine has provided its translation. Naturally, these three phases of processing can be combined in several ways.



# Machine Aided Human Translation

- MAHT: Helps the human translator whenever he or she asks for it, in one way or the other.
- In its simplest form, MAHT-software can be spell check and grammar check.
- The slightly more advanced MAHT-software includes electronic dictionaries, terminology databases etc.
- Today, the most advanced form of MAHT-software is Translation Memories (TM). This also includes “Computer Aided Translation” or CAT.



# Machine Translation for African Languages

- OBTranslate: Aims to translate more than 2,000 African languages.
  - According to its creator, 63 per cent of the population in Sub-Saharan Africa do not have access to global markets because of language barriers.
  - Machine Learning & Computer Aided Translation
  - Translate (textual), transcribe (from video/audio to text), and create subtitles (translate videos with sub-titles)

<https://www.scidev.net/global/communication/news/digital-tool-aims-to-translate-2-000-african-languages-1x.html>

- Google Translate: English, Afrikaans, isiZulu, isiXhosa and Southern Sotho



# Machine Translation for African Languages

According to L. Martinus and J.Z. Abbott (2019),

- African languages are numerous, complex and low-resourced;
- Datasets required for machine translation are difficult to discover; and
- Existing research is hard to reproduce.
- Minimal attention has been given to machine translation for African languages so there is scant research regarding the problems that arise when using machine translation techniques.
- Trained models to translate English to five of the official South African languages (Afrikaans, isiZulu, Northern Sotho, Setswana, Xitsonga), making use of modern neural machine translation techniques.
- Results obtained show the promise of using neural machine translation techniques for African languages.



# Machine Translation for African Languages

- Development of an automated English-to-local-language translator as a model to bridge the communication gap in Nigeria and other multilingual settings.
- Scope of implementation: English-to-Efik language translation.
  - A bilingual dictionary containing 500 words and 25 corpuses was designed to provide optimum direct translations regarding the Efik language.
  - The quality of translation is influenced by the *size of bilingual dictionary* provided and *accuracy of generative rules* declared in the system's shell.
  - The software successfully translates English text to Efik using the words and corpuses available in a bilingual dictionary.

[https://www.researchgate.net/publication/326722662\\_Development\\_of\\_an\\_automated\\_English-to-local-language\\_translator\\_using\\_Natural\\_Language\\_Processing](https://www.researchgate.net/publication/326722662_Development_of_an_automated_English-to-local-language_translator_using_Natural_Language_Processing)



# Translation Tools vs Speech Technology

- Translation Tools: Translate a written text from one natural language (the source language) into a text in another natural language (the target language).
- Observation:
  - Translation tools are developed with the purpose of translating written texts between various languages, in contrast to, for example, speech recognition systems. Speech recognition systems typically “translate” within the same language from one medium (speech) to another (text).
  - It is a case of the translation of texts from a *natural language* > Natural language processing (NLP) framework.
  - Other sensitivities: Digital/language proficiency; disability.



Thank you for your attention.

