Multimodal, Multilingual Dynamic Stories for Literacy Development and Language Learning

Laurette Marais
Voice Computing
NextGen Enterprises and Institutions
CSIR
Pretoria, South Africa
lmarais@csir.co.za

Ilana Wilken
Voice Computing
NextGen Enterprises and Institutions
CSIR
Pretoria, South Africa
iwilken@csir.co.za

Laurette Pretorius
Division of Computer Science,
Department of Mathematical Sciences,
Stellenbosch University
Stellenbosch, South Africa
lpretorius@sun.ac.za

Lionel Posthumus
CALT@UJ
University of Johannesburg
Johannesburg, South Africa
lionelp@uj.ac.za

Figure 1: Example Game task of the *Ngiyaqonda!* application. Users listen to a prompt sentence and have to select the correct tokens to write the sentence themselves.

CUI '23, July 19–21, 2023, Eindhoven, Netherlands

ACM ISBN 979-8-4007-0014-9/23/07.
https://doi.org/10.1145/3571884.3604303

ABSTRACT

Early literacy acquisition has been shown to be life-changing. In resource-scarce multilingual, multicultural societies in which appropriately skilled human resources are scarce and the acquisition of early literacy is a huge challenge, the innovative and creative use of technology offers new ways of tackling this strategic educational challenge. Speech technology is particularly suited to this problem, since children can speak long before they can read and write. In this work we present the *Ngiyaqonda!* project, in which a multimodal learning environment is being developed to facilitate both literacy development and language learning, with the ultimate aim of solving a unique and pressing challenge facing South African foundation phase learners.

CCS CONCEPTS

• **Applied computing** → **E-learning**; **Interactive learning environments**; • **Computing methodologies** → **Natural language generation**; **Speech recognition**; *Phonology / morphology*.

KEYWORDS

text-to-speech, literacy, language learning, resource-scarce languages

ACM Reference Format:

Laurette Marais, Ilana Wilken, Laurette Pretorius, and Lionel Posthumus. 2023. Multimodal, Multilingual Dynamic Stories for Literacy Development and Language Learning. In *ACM conference on Conversational User Interfaces (CUI '23), July 19–21, 2023, Eindhoven, Netherlands.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3571884.3604303

1 INTRODUCTION

During the first three years of their education, South African children are taught to read in one of the eleven official languages of South Africa, and for the vast majority, this means being taught in their home language. However, by the end of the foundation phase (Grades 1-3), across all languages, 78% of learners cannot read for comprehension in their home language[6]. The reading deficit is particularly acute in the case of learners who are taught through an African language as language of learning and teaching (LOLT) during the foundation phase. In the case of isiZulu, the number of learners who cannot read for comprehension is 87% and for Sesotho sa Leboa, that number is 93%. That means that an overwhelming number of South African foundation phase learners "cannot read for meaning or retrieve basic information from the text to answer simplistic questions"[6]. The reasons for this relate broadly to the inability of educators to teach reading successfully[11] , and in particular, to teach the fundamentals of reading in the African languages[9, 18, 19].

When learners enter the intermediate phase in Grade 4, the LOLT shifts to English, which means that children who were taught to read and write in, for example isiZulu and Sesotho sa Leboa, must be able to read and write English in almost all the subjects they are taught. Learners and educators are essentially forced to "play catch-up" in the intermediate phase[10].

> For these learners this sudden shift from Home Language to English FAL (first additional language) as the medium of instruction is doubly challenging, as they do not have the language proficiency advantage in the FAL that home language speakers have, plus they do not have grade appropriate reading skills in their Home Language to support a gear-shift change in reading in a FAL[12].

Language and speech technology is uniquely positioned to help address the problem, especially if it has been developed first and foremost to be reliable. Since foundation phase learners have already developed a level of oral competency in their home language, this can be leveraged to support the development written competency - first in their home language, and eventually also in the language in which they will be taught from Grade 4 onward.

Digital technology is inherently scalable in at least two pertinent ways. Firstly, a mobile application that has been shown to support learning objectives without immediate direction by a teacher can be deployed to support as many children as can be supplied with appropriate devices. This can alleviate human resource scarceness. For now, however, the application is aimed at providing support for foundation phase teachers.

Secondly and more importantly, language technology, and especially natural language generation in conjunction with text-to-speech, can be employed to supply a possibly infinite number of learning examples. At the very least, even with a limited, carefully chosen vocabulary, the ability to generate thousands of examples is easily achieved via a computational grammar. This would allow an application to provide learners with varied content every time they use the application, while appropriate limitation of the domain will ensure that the intended learning objectives are achieved.

2 CORE TECHNOLOGIES

The *Ngiyaqonda!* project aims to develop a mobile application that employs grammar-based natural language generation and speech technology to create a multimodal, multilingual learning environment that provides learners the opportunity to practise reading and sentence composition. It is aimed at learners in Grade 3, who are assumed to have mastered basic decoding[16] and would benefit from guided repetition. Initially, the application focuses on the learner's home language, before providing a gradual introduction of the target language, which would typically be the LOLT of their intermediate phase education. In this section, we discuss the core technologies that form the basis of the learning environment.

2.1 Grammar-based NLG

Natural language generation (NLG) is a task concerned with generating natural language from machine readable representations of data. The framework employed in this work is Grammatical Framework (GF), "a special-purpose programming language for writing grammars."[14] The GF system includes a compiler[3] which compiles GF code into Portable Grammar Format (PGF) files, as well as a runtime system for interacting with PGFs. The C runtime system has bindings to a number of other languages, including Java and Python[4, 15].

A primary focus of GF has always been on the development of so-called *application grammars*. Simply put, application grammars are grammars that model utterances within a specific domain. GF is inherently focused on multilingualism: at the core of each grammar is an abstract syntax defining abstract categories and functions. In an application grammar, these categories and functions are semantic in nature, defining the kinds of entities or concepts in the domain, and how they can be combined to form utterances. In other words, the categories and functions can be applied to produce a (possibly infinite) set of abstract syntax trees, which represent all utterances of the domain. The categories and functions are mirrored in a set of language specific concrete syntaxes, which define so-called linearisation categories and functions. Linearisation is the process by which the GF runtime produces strings from abstract trees ; conversely, parsing is a non-trivial inversion of linearisation[13], whereby the GF runtime produces trees from strings defined by a concrete syntax of the grammar. Putting it in NLG terms, during linearisation, the GF runtime receives as input machine readable data in the form of a tree, and produces natural language as output. Multilingual NLG is achieved by linearising the same abstract syntax tree (which represents an utterance language independently) using multiple concrete syntaxes, thereby producing language specific utterances that express the same meaning.

An application grammar, especially a multilingual application grammar which is aligned at the semantic level, is an ideal choice of NLG system for providing practise examples for reading and sentence composition. It allows control of the semantic and linguistic complexity of the utterances of the domain, and crucially, provides complete control over semantic and grammatical correctness, which in the case of isiZulu, as agglutinating language, places a particular focus on the correctness of the morphology and syntax. Learners who are still in the process of cementing their reading, writing and language skills cannot be exposed to technology that purports to support acquisition of these skills if the technology cannot be relied upon for correctness.

The set of grammars that have been developed for piloting the application are based on a story published by the VulaBula project[1], which provides open source education materials in all 11 official languages of South Africa.

An analysis of types of suitable sentences found in the isiZulu and English versions of the story was done, from which a language independent abstract syntax was derived. So far, concrete syntaxes for isiZulu and English have been developed[1] - Afrikaans and Sesotho sa Leboa concrete grammars are currently under development. This means that each abstract syntax tree defined in an abstract module can be linearised as contextually appropriate isiZulu and English sentences.

The set of grammars comprises four different abstract syntax modules that represent progression in semantic complexity. As the grammars become more semantically complex, the linguistic complexity similarly increases: different kinds of noun modification, different kinds of adverbial phrases (verb modification), and the use of infinitives are introduced. A limited vocabulary is employed, covering 11 verbs, 28 nouns, 7 adjectives (in the case of isiZulu, this includes both adjective and relative stems) and 5 adverbial phrases.

The power of computational grammars for generating focused but diverse practice examples is especially apparent when one compares the sizes of utterance sets defined by the first and the fourth grammars. The first grammar defines 216 utterances using a subset of the vocabulary, while the fourth grammar, utilising the full (albeit small) vocabulary mentioned above, defines 45 905 utterances. It should be emphasised that each of these utterances conforms to the semantic constraints of the abstract syntax grammar, and are therefore contextually plausible. Moreover, each of them can be rendered as a contextually appropriate isiZulu and an English sentence (and soon also in Afrikaans and Sesotho sa Leboa). The GF application grammars therefore succeed in providing a high level of pedagogical control, while also providing a practically limitless supply of multilingual practice sentences for the mobile application.

[1]This has been facilitated by the available GF resource grammars for isiZulu and English[2]

2.2 Text-to-speech

Given the scope of the text-based NLG component and the need for extensibility in future, the only viable option for supplying the vital speech component to the application is text-to-speech (TTS) technology. For this purpose, the Qfrency[7, 8] TTS system is employed. Qfrency is a suite of synthetic voices, supporting all of the 11 official South African languages. For English, isiZulu, Sesotho sa Leboa and Afrikaans in particular, both male and female adult voices are available.

The requirements of TTS within a multimodal learning environment, where the learning focus is on literacy and language learning, is that it is highly reliable in terms of intelligibility and with a high level of naturalness. The Qfrency voices that are DNN-based meet these criteria. Qfrency also supports the Speech Synthesis Markup Language, which provides additional control over, inter alia, pronunciation. While no such words have yet been identified, it is possible that proper nouns, typically character names, may present a problem for a language-specific TTS system. Accepting SSML allows the client application to supply additional information that directs the TTS system to render the correct pronunciation.

3 THE *NGIYAQONDA!* APPLICATION

The mobile application is implemented as an Android application, since Android-based mobile phones dominate the South African market with a market share of over 80%[17]. It provides the user with a progressive learning journey based on the user's selected home language and target language. The application is split into a backend, which is responsible for the structuring and presentation of all language and speech content, and a lightweight frontend that provides navigation from a user login process to lesson and task selection screens, where lessons consist of a related collection of tasks.

The backend is responsible for serving a lesson and task structure, as well as parameters such as the number of sentences presented in the task, a minimum pass score where applicable, the number of alternative words to provide for selection, etc. Each of the eight lessons is supported by one of the four multilingual GF grammars, which facilitates the progression in complexity as the user completes lessons. Tasks can be monolingual or multilingual. Initially, the user is only presented with monolingual tasks in their home language. Eventually, the target language is introduced with the requirement that the user exhibit passive knowledge of the target language to make progress. Finally, the multilingual tasks require active knowledge of the target language.

There are three kinds of tasks, each presented as a separate activity screen, namely Story tasks, Write tasks and Game tasks. For the sake of brevity, we briefly mention the Story task, and provide a more detailed descriptions of the Write and Game tasks.

The Story task aims to introduce the user to the story on which the content of the application is based, as well as introducing the central touch mechanism for sentence construction, namely dragging and dropping from and to specific areas on the screen.

The Write task (shown in Figure 2) provides the user with the opportunity to practice reading and sentence composition in a freestyle way, while familiarising the user with the growing complexity of utterances in the domain. Multiple words are presented in the *selection area*, from which the user can drag words to the *authoring area* in order to compose sentences. The grammar ensures that only grammatically and semantically correct words are presented. This kind of interface, driven by a GF grammar, has long been called the "fridge poetry" interface, mimicking the way one composes sentences using fridge magnets[5]. The crucial difference, of course, is that the grammar ensures that sentences are semantically and grammatically well-formed. In this way, users are able to compose any sentence allowed by the grammar, allowing for free experimentation within the domain. Dedicated buttons allow users to undo their selections, as well as to "write" completed sentences to the *authored area*. In multilingual versions of the task, the sentences appear as pairs in the authored area, namely both in the home language and in the target language. The lesson journey requires that users complete tasks consisting of composing sentences in their home language first, before later requiring that they compose sentences in their target language. Authored sentences are audio enabled via TTS as soon as they appear in the authored area. A user completes the task when the predefined number of sentences has been composed and listened to.

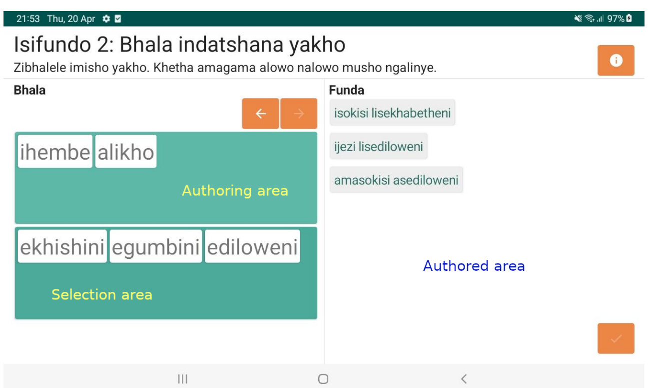

Figure 2: Example Write task of the *Ngiyaqonda!* application

The Game task (shown in Figure 1) exploits the existing oral competency in the learner's home language, and developing oral competency in their target language, to gamify reading and sentence composition. In these tasks, a number of random prompts from the grammar are generated and presented in turn to the user via a listening button. The user listens to the current prompt, and has to construct the corresponding sentence by selecting, word for word, the correct token from the selection area. The available words are updated at each selection, ensuring a semantically and grammatically plausible set of options each time. Once the entire prompt sentence has been composed, the sentence is moved to the authored area automatically as audio enabled text, and the game continues with the next prompt.

There are three configurations of the game. In monolingual tasks, the prompt language and the composition language are the same, namely the user's home language. Then, the target language is introduced by providing the prompts in the target language and requiring the user to compose the correct translation of the prompt in their home language. This requires passive knowledge of the target language. Finally, the prompt is provided in the user's home language and they are required to compose the translation in their target language, which requires active knowledge of the target language.

4 PLANNED PILOTING

The application will be piloted in the third quarter of 2023. A school where isiZulu is the LOLT for Grades 1 to 3 and English the LOLT from Grade 4 onward will serve as pilot site. With the assistance of an assigned Grade 3 teacher, a pilot group and control group of 10 learners each will be identified. Existing assessments, including an EGRA test in the home language, carried out by the school, will be utilised to ensure that the groups are comparable in terms of academic ability.

The pilot will consist of the pilot group using the application for 30 minutes per week for eight weeks during a dedicated reading period. All activity will be logged against each user in an anonymised way. This will provide general insights into use of the application by the group of learners. After the piloting period, home language EGRA tests will again be conducted on both groups, in which we hope to observe a positive effect from exposure to the application. The results of these tests, as well as interviews with relevant teachers, will form the basis of the assessment of the efficacy of the application to improve reading and sentence composition in learners' home language and target language. After the piloting period, all Grade 3 learners at the school, including the control group, will be provided with the opportunity to use the application.

5 CONCLUSION

We have presented a multimodal, multilingual mobile application, based on a children's story, which aims to support literacy development and language learning within the South African context. The environment mimics the inherent multimodality of human language, mirroring its oral and written forms via TTS and multilingual NLG, in order to create a dynamic learning environment for South African foundation phase learners. The use of GF application grammars provide a highly reliable, controlled language domain within which learners are exposed to semantically and grammatically well-formed sentences. While the application currently focuses on individual sentences within the domain in order to support literacy development through practising reading and sentence composition, future work will investigate the incorporation of discourse level NLG, in which automatic comprehension testing, the ability to practise narrative composition and the opportunity to engage in interactive dialogue in multiple languages will be made possible. This will serve to further support home language literacy development, as well as the shift to English that a large number of South African learners must make in the intermediate phase of their education.

ACKNOWLEDGMENTS

The *Ngiyaqonda!* project is sponsored by the South African Department of Sport, Arts and Culture.

REFERENCES

[1] 2018. Vula Bula. https://vulabula.molteno.co.za/ Accessed on April 20, 2023.
[2] 2023. GF Resource Grammar Library (RGL). https://github.com/GrammaticalFramework/gf-rgl Accessed on April 20, 2023.
[3] Krasimir Angelov. 2009. *GF Runtime System.* Ph. D. Dissertation. University of Gothenburg.
[4] Krasimir Angelov. 2017. Using the Python binding to the C runtime. http://www.grammaticalframework.org/doc/runtime-api.html#python Accessed on April 20, 2023.
[5] Björn Bringert, Krasimir Angelov, and Aarne Ranta. 2009. Grammatical framework web service. In *Proceedings of the Demonstrations Session at EACL 2009.* 9–12.
[6] Sarah J Howie, Celeste Combrinck, Karen Roux, Mishack Tshele, Gabriel Mokoena, Nelladee McLeod Palane, et al. 2017. *PIRLS literacy 2016: South African highlights report (Grade 4).* Technical Report. Centre for Evaluation and Assessment (CEA).
[7] Johannes A Louw. 2023. Qfrency. http://www.qfrency.com/ Accessed on April 20, 2023.
[8] Johannes A Louw, Avashlin Moodley, and Avashna Govender. 2016. The Speect text-to-speech entry for the Blizzard Challenge 2016. (2016).
[9] Taimi-Ndapandula Nghikembua. 2020. Learning to read and spell in Oshikwanyama language: Precursors, dynamics and teacher knowledge of early literacy instruction. *JYU dissertations* (2020).
[10] Elizabeth J Pretorius. 2014. Supporting transition or playing catch-up in Grade 4? Implications for standards in education and training. *Perspectives in Education* 32, 1 (2014), 51–76.
[11] Elizabeth J Pretorius and Nanda M Klapwijk. 2016. Reading comprehension in South African schools: Are teachers getting it, and getting it right? *Per Linguam: a Journal of Language Learning= Per Linguam: Tydskrif vir Taalaanleer* 32, 1 (2016), 1–20.
[12] Elizabeth J Pretorius and Lieke Stoffelsma. 2021. The potential for word growth in Grade 3 in disadvantaged urban schools: Factors that facilitate or impede the transition to Grade 4. *Multilingual classroom contexts: Transitions and transactions* (2021), 17.
[13] Aarne Ranta. 2006. Type Theory and Universal Grammar. *Philosophia Scientae* CS 6 (2006), 115–131. https://doi.org/10.4000/philosophiascientiae.415
[14] Aarne Ranta. 2011. *Grammatical Framework: Programming with Multilingual Grammars.* CSLI, Stanford, California.
[15] Aarne Ranta, Krasimir Angelov, Normunds Gruzitis, and Prasanth Kolachina. 2020. Abstract syntax as interlingua: Scaling up the grammatical framework from controlled languages to robust pipelines. *Computational Linguistics* 46, 2 (2020), 425–486. https://doi.org/10.1162/COLI_a_00378
[16] Kathleen Rastle, Clare Lally, Matthew H Davis, and JSH Taylor. 2021. The dramatic impact of explicit instruction on learning to read in a new writing system. *Psychological Science* 32, 4 (2021), 471–484.
[17] StatCounter. 2023. Mobile Operating System Market Share South Africa. https://gs.statcounter.com/os-market-share/mobile/south-africa Accessed on April 20, 2023.
[18] Rosemary Wildsmith-Cromarty and Robert J Balfour. 2019. Language learning and teaching in South African primary schools. *Language Teaching* 52, 3 (2019), 296–317.
[19] Carien Wilsenach. 2019. Phonological awareness and reading in Northern Sotho–Understanding the contribution of phonemes and syllables in Grade 3 reading attainment. *South African Journal of Childhood Education* 9, 1 (2019), 1–10.