Automating question generation and marking of language learning exercises for isiZulu

Nikhil GILBERT[a] and C. Maria KEET[a,1]
[a] *University of Cape Town*

Abstract. Increase in isiZulu language learning is hampered by the predominantly manual approach to creating and marking homework and test exercises. Extant computer-assisted language learning platforms cannot handle the intricacies of agglutination in isiZulu and related languages. We seek to address this by designing a controlled natural language-based exercise generator and marker for isiZulu. This consists of question and answer sentence templates for exercise types, reusable algorithm snippets as grammar library, a small corpus of words and sentences to be used by the system, a constrained sentence generator to combine the right type of words, and finally the exercise creation and automated marking system. The preliminary evaluation shows encouraging results.

Keywords. Language learning; IsiZulu; Controlled Natural Language; Computer-Assisted Language Learning; Corpus

1. Introduction

Africa is culturally a diverse continent with hundreds of languages. Interactions between various African cultures throughout history have enriched the languages with various nuances and similarities [11]. South Africa has 11 official languages [20] and many residents are bilingual or multilingual, either due to home circumstances or through learning another language at school [5]. IsiZulu is the most widely spoken language of South Africa with 23% of the population speaking it as a first/home language [14] and about half of the population speaks it to some degree. Recently, the University of KwaZulu-Natal has introduced a compulsory introductory course in isiZulu for all its first-year students, out of a total enrollment of about 45000 students (i.e., roughly 9000 students per year). Other universities require this for selected degree programmes only, also for a very closely related language, such as isiXhosa, amounting to some 500-1000 students per year per university in South Africa.

There are only very limited learning resources for isiZulu, however, which are predominantly classroom and paper-based. The lack of computational resources for learning isiZulu resulted in the curriculum having been developed around teaching isiZulu using outmoded methods to assess students [5]. This also makes it difficult to find a means of widely distributing any new developments of language learning techniques. The manual marking of language learning exercises for such large amounts

[1] C. Maria Keet, Department of Computer Science, University of Cape Town, South Africa; E-mail: mkeet@cs.uct.ac.za.

of students has shown to cause many problems. The principal issues are: 1) prone to errors in marking, 2) loss of scripts, 3) time taken to return the work to students, and 4) limited options to assess the students' progression in language learning [18, 25]. In addition, many of the currently existing isiZulu corpora are dated or do not provide adequate material for properly studying a language [16, 28]. Automation may alleviate these issues. This requires a system that 1) controls, or limits, the language appropriately for the level of language learning, and 2) provides structured exercises that can be marked automatically. Research and tools for computer-assisted language learning (CALL) exercises focus more on either just vocabulary and sentence grammar [6, 7], advanced tasks such as reading comprehension, essay writing, and meta-cognition of the grammar [3, 8, 31], or crowd-sourced user input, such as on Duolingo. A common characteristic of the NLP-mediated approaches is that those languages are well-resourced, such as having POS taggers, parsers, analyzers, and grammars for sentence construction, to generate questions and answers for automated marking. Reuse of such existing tools is infeasible, not just because isiZulu is still underresourced, but because isiZulu is a Bantu language characterized by agglutination and its characteristic noun class system (NCS). That is, it has a very different morphology and grammar from Indo-European languages and it is the NCS that causes the steep learning curve for beginners: one needs to know the NCS even before looking up words in the dictionary.

We seek to address these issues by developing a controlled natural language (CNL)-based question generation and automated marking system that is inspired by pen and paper-based exercises used by teachers to date. This is realized by, mainly, embedding extant computational models of isiZulu grammar and morphology and repurposing isiZulu natural language generation algorithms of [2,15] to compute the question answers. The modular approach to question and answer template specification can enable teachers to easily assess students with varied exercises. The current system can generate almost 40000 unique question sentences, thanks to the templates and words taken from a newly designed small relevant corpus. From a language learning viewpoint, the questions principally address the issue of form exposure and practice, which, given the centrality of the NCS, is crucial for advancing to an intermediate level.

The remainder of the paper is structured as follows. We summarize some basic aspects of isiZulu and related works on CALL systems in Section 2. Section 3 describes the design of the CNL-based CALL system. Section 4 contains an evaluation. We reflect on the system in Section 5 and conclude in Section 6.

2. Background and Related Work

We first describe salient aspects of isiZulu, which helps contextualizing the related work on CALL systems and the requirements for them.

2.1. Some aspects of (Computational) Linguistics for isiZulu

There are a few algorithms for generating isiZulu words or text, being pluralisation of the isiZulu noun [2] and a CNL for ontologies that also includes conjugation [15]. Pluralising nouns is based on the noun class (NC) that a noun is classified in; e.g., the singular *indlovu* 'elephant' in NC9 is pluralized as *izindlovu* in NC10. The 17 isiZulu noun classes are listed by singular/plural pair in Table 1, from which rules were developed that also include various deviations and exceptions. Due to sameness of prefixes for some NCs, the pluralizer uses both a morphological and semantic approach, through analyzing the first few letters of a noun in combination with the NC stored with the noun (not doing so reduces the accuracy to a mere 50% [2]). This also hints at a difficulty in isiZulu language learning, which is to figure out which noun class the noun is of, especially in those cases where the prefix is the same for the noun in the singular but different for the plural or vice versa; e.g., NC1 and NC3 have the same prefix *um(u)-* and NC10 can have either NC9 or NC11 as singular, which have different prefixes. There are also nouns that have the same stem but a different prefix and therewith obtain different meanings; e.g., *umuntu* 'human' and *ubuntu* 'humanity'.

After the NCS, the next main step at the introductory level is conjugation of the verb, which is governed by the NC of the subject by means of the so-called subject concord (SC), among other possible constituents that can be seen as a slot system [14]. Each NC has its own SC. For instance, with the concordial agreement underlined, *umuntu uhamba* 'the human goes' and, in the plural, *abantu bahamba*, but if, say, the elephant goes, it is *indlovu ihamba* (plural *izindlovu zihamba*). The negation is governed by the negative SC, which is also specific to each NC; e.g., *umuntu akahamba* 'the human does not go' and *indlovu ayihamba*. There are a few cases where there is the same negative SC for different NCs, which ups the exercise difficulty level; e.g., NC4 and NC9 both take *ayi-* [14]. Algorithmically, once the NC of the noun is fetched, adding the SC to the verb uses a NC-driven lookup table and checks for phonological conditioning required for the few vowel-commencing verb roots.

The verb can take various extensions, such as certain prepositions that are inserted in the verb and the wh-questions that always go at the end [14]. For instance, 'to work for' *ukusebenzela* (*uku-* 'to', *-sebenz-* verb root of 'work', *-el-* is the reciprocal extension, *-a* the final vowel) and *uvelaphi?* 'where [do you/does (s)he] come from?'. The NCS further governs prefixes of, among others, adjectives and possessives into a comprehensive concordial agreement system that is taught at an intermediate level. Therefore, an excellent command of the NCS is of paramount importance to advance in isiZulu and related languages that use a NCS.

While the multitude of all components are still being investigated, it is clear that this lends itself well for a modular building up of a CNL with a set of templates of varying complexity for learning exercises, where the teacher would be able to select which grammatical components should be included as well as which noun classes.

NC	Prefix	Examples	NC	Prefix	Examples
1	um(u)	*umuntu* 'human'	9a	i	*ivazi* 'vase'
2	aba	*abantu*	(6)	ama	*amavazi*
1a	u	*ugogo* 'grandmother'	9	i(N)	*indlovu* 'elephant'
2a	o	*ogogo*	10	izi(N)	*izindlovu*
3a	u	*ushizi* 'cheese'	11	u(lu)	*uphawu* 'mark'
(2a)	o	*oshizi*	(10)	izi(N)	*izimphawu*
3	um(u)	*umfoloko* 'fork'	14	ubu	*ubuhle* 'beauty'
4	imi	*imifoloko*			
5	i(li)	*igama* 'name'	15	uku	*ukuhamba* 'to go'
6	ama	*amagama*			
7	isi	*isilwane* 'animal'	17	ku	(locatives)
8	izi	*izilwane*			

Table 1. isiZulu noun classes with prefixes for each noun class with examples (based on [2]); the odd-numbered classes are the singular and even numbers are their plural and the 'N' denotes 'n', 'm', or ''.

2.2. Computer-Assisted Language Learning

Principles of integrative CALL are described in [17], integrating reading, writing, speech, and listening to deliver lessons. Random generation of questions is an important part of CALL systems, as repeated texts suffer from a phenomenon known as the “practice effect” [23] when the same material is presented to a user repeatedly.

There are typically three components of a CALL system: a grammar rule set [26] and sentence generator, the contextually relevant corpus [10, 17], and the language learning exercises that may have hints and answers [26]. The rule set of the language’s grammar has to be established first, as the number of language learning exercises may be constrained by it [26]. The grammar rules need to consist of, at least, a series of computationally modelled parts of speech (POS), especially the noun and verb [4]. With a ‘slot system’ to add and remove components, a learning platform would then be only limited to a teacher’s ingenuity.

The language learning exercise question set should be independent of the rule base so as to avoid repetition in the code and avoid confusion in future parallel development [13]. The questions need to challenge learners also in the basic and mandatory contexts [1], yet also respect the scope of pen-and-paper exercises, such as [29] for isiZulu.

The variability in questions normally comes from a sufficiently large corpus from which to draw words and sentence fragments. It is important that the corpus is contextually relevant and contains words of modern communication that can be used in every day speech [17, 27]. The corpus has to be annotated with the correct POS tags, which is subsequently used by the grammar rules [21]. This is a non-trivial step for isiZulu: the Ukwabelana corpus is tagged with a gold standard [28] but contains outdated texts [22], whereas the news item corpus with recent texts [16] is not POS tagged and there is no fully functional isiZulu POS tagger.

Natural language generation has been used in CALL systems. Closest to our scope, there is a French grammar and question bank for fill-in-the-blank questions and sentence shuffle [24], and question/answer sentence pairs with SemTAG (a Feature-Based Lexicalised Tree Adjoining Grammar) and transformation rules for whole sentences [7]. NLG also has been used to assist with essay writing [8]. There are more distantly related efforts, such as CNL-based question generation from texts for grammar concept questions [3] and reading understanding questions [31]. These types of questions are at a much more advanced level of language learning than the contents of currently available textbooks for isiZulu. Also, the demand for a computational approach for those type of questions is not nearly as pressing as the questions for thousands of students at the introductory level of language learning.

Currently, very few CALL systems for isiZulu (and Bantu languages more generally) are available that use an integrative learning strategy [12, 19]. We thus also consulted current paper-based systems [29] to transfer into automated templates.

3. Specification and Design

The platform was designed with a layered architecture, with each layer feeding either information, or services to the next, as can be seen in Fig. 1. A layered architecture is claimed to have the benefits of ease of scalability and the convenience of concurrent development [9]. The underlying idea of the present design is similar to [7] in the sense of source and target sentences and transformations, but it is specified as a CNL with word component exercises and it uses algorithms to obtain the transformations, rather than tree transformations with a grammar (a formalisation of sentence grammar does not exist for isiZulu). We discuss each component in the following sections.

3.1. Grammar Rule Sets

Because isiZulu is an underresourced language, any resource developed should be easy to reuse in other systems. Therefore, the system has a separate support library for isiZulu grammar, which also includes support functions such as the randomized sentence generator. This ensures that no recoding is required for each new CALL exercise, as the set of methods and classes allow flexible reconfiguration of the constituents of the words and sentences.

The grammar rule set of the system consists of the following components, (harmonized into Java): a new singularizer by 'reversing' the pluraliser algorithm; the pluralizer of [2]; the verb grammar represented as a CFG [14] with a substantial new set of verb roots; the algorithm for verb conjugation from a CNL for ontologies [15]. These are sufficient to create basic sentences for the exercises and their answers (see below). Further, each morpheme can be obtained so that exercises may be constructed with individual grammar components only as well; e.g., getVerbGrammar("SC") will return a list of all subject concords. There are further functions, notably to generate a specific verb form, which allows the user to stipulate the morphological specificities of the verb they would like to generate, and to check consistency of the verb form with the CFG. For instance, if a verb has both a positive and negative subject concord, it violates the CFG, so the function will then only append affixes consistent with the positive subject concord. Current algorithms for isiZulu sentence generation are not perfect [15], which is largely due to the pluralizer that hovers between 92-100% accuracy depending on the test set [2].

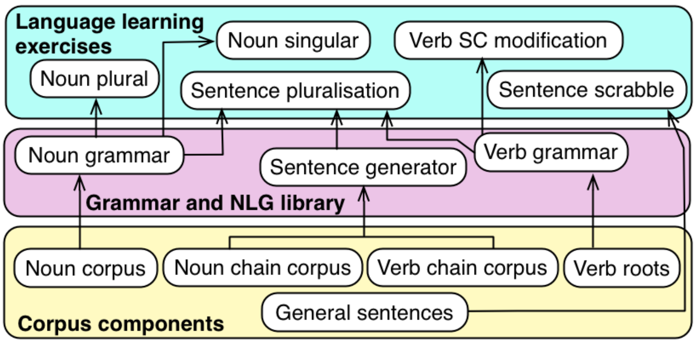

Figure 1. Architecture of the CNL-driven CALL system. The arrows indicate which upper layer components make use of the lower layer components.

3.2. Corpus Specifications

The two existing isiZulu corpora turned out to be unsuitable for the task. Ukwabelana [28] was contextually inadequate for it consists of the bible and novels, it contains outdated isiZulu, or did not contain words that were simple enough for novice language learning. The portion of the isiZulu National Corpus (INC) [16] turned out to be too variant with the number of words. We therefore constructed a small corpus consisting of five lists of manually POS-tagged words and sentences, adhering to consistency in the style of annotation with other corpora. The corpus' content was also inspired from a beginner's language learning book in order to use words that were deemed most relevant for conversational context [29]. The corpus includes: a list of common nouns (*n*=231) and verbs (*n*=59) typical for language learning, such as *umfundi* 'learner', *ikhaya* 'home', *-enza* 'do', *-hamba* 'go', *thenga* 'like'; common phrases spoken in isiZulu (*n*=60); a morphologically analysed and tagged word list (*n*=10040); and the noun and verb chain lists for sentence generation that is described in the next section.

3.3. Sentence Generation Algorithm

We created a sentence generation algorithm that draws from the annotated corpus. The possible sentence structures supported in the system are either of the high-level pattern "`<noun> <verb>`" or "`<noun> <verb> <noun>`". In order to string the right words together, it uses two chained lists of words. The Noun Chain List is a list of 231 nouns, each of which has a list of one or more suitable verbs that can follow it; e.g., animals can 'eat' and 'drink' but a radio cannot. The generator will select a noun at random from the list, and then select verb at random from that chain. For each verb in Verb Chain List, there is a list of suitable nouns that also can be selected at random. It continues to follow these chains until a sentence with the specified structure is built. The list is annotated with whether the verb needs an object or not, as the latter can be used in the "`<noun> <verb>`" pattern, but the former cannot (indicated with <t>). For instance, given the examples in Figure 2, the algorithm selects *ubaba* (NC1), and picks *washa* 'wash' from the list of verbs in the Noun Chain List. Then in the Verb Chain List, *washa* lists what objects can be washed, such as *imoto* 'car' and thus can be used in the "`<noun> <verb> <noun>`" pattern (but need not).

Noun chain list	Verb chain list
ubaba <1a> washa;sula;faka;khuluma	washa <t> imoto;umshini;umnyango
umzali <1;s> ALL_v;e_dumisa;e_cisha	sula <> ifasitela;imoto;ipuleti
	khuluma <t> ALL_1;ALL_1a

Figure 2. Illustration of the Noun Chain and Verb Chain lists options: nouns may take some verbs, all verbs "ALL_v" with or without exceptions "e_", and verbs may go with some specified classes (e.g., "ALL_1") or with nouns in specific noun classes only, such as all people (NC1, NC1a) 'speak' *-khuluma*.

3.4. Language Learning Exercises

The language learning exercises are built using the corpus and grammar library and are alike existing pen-and-paper based isiZulu language learning challenges [29] and it draws inspiration from Duolingo's sentence scrabble [30]. The types of exercises include reordering scrabbled sentences, pluralizing and singularizing nouns with the correct prefixes, and modifying the verb's SC in a sentence based on the noun as well as changing positive/negative that requires learners to change the morpheme of the verb to be in agreement with the correct verb grammar rules. Figure 3 illustrates some of these exercises with questions and computed (correct) answers. For instance, questions of type 1 and 2 help the learner to spot wrong concordial agreement, provide a hint as to what the plural might be, and assist in discovering sound agreement between the noun prefixes and concords (e.g., *aba- ba-*, *isi- si-*).

1. Pluralise the subject:
Q: Umfowethu bayaphuza
A: Abafowethu bayaphuza
Q: Indlovu zidla ihlamvana
A: Izindlovu zidla ihlamvana
2. Pluralise the verb:
Q: Oyihlo utheleka
A: Oyihlo batheleka
3. Pluralise the sentence:
Q: Isitshulu simnandi
A: Izitshulu zimnandi
Q: Umfowethu usula inkomishi
A: Abafowethu basula izinkomishi
4. Positive/negative subject concord:
Q: Batotoba
A: Abatotobi

Figure 3. Illustration of sample exercises (questions and answers) the system is able to produce.

Regarding the templates, recall that the two sentence patterns are "<noun> <verb>" and "<noun> <verb> <noun>". The actual exercises templates have <noun> constructed from prefix[SG/PL] + stem and the verb is composed of [Negative]Subject Concord + VerbRoot + [Negative]FinalVowel, taking into account phonological conditioning in the agglutination. The examples in Figure 3 are based on the following template pairs, with the change for the answer underlined and the wrong concordial agreement indicated in italics in the question, where the answer is computed with the algorithms mentioned in Section 3.1:

1. Q: <*prefixSG*+stem> <*PLSC*+VerbRoot+FV>
   A: <prefixPL+stem> <PLSC+VerbRoot+FV>
   Q: <*prefixSG*+stem> <*PLSC*+VerbRoot+FV> <prefixSG+stem>
   A: <prefixPL+stem> <PLSC+VerbRoot+FV> <prefixSG+stem>

2. Q: <*prefixPL*+stem> <*SGSC*+VerbRoot+FV>
   A: <prefixPL+stem> <PLSC+VerbRoot+FV>

3. Q: <prefixSG+stem> <SGSC+VerbRoot+FV>
   A: <prefixPL+stem> <PLSC+VerbRoot+FV>
   Q: <prefixSG+stem> <SGSC+VerbRoot+FV> <prefixSG+stem>
   A: <prefixPL+stem> <PLSC+VerbRoot+FV> <prefixPL+stem>

4. Q: <PLSC+VerbRoot+FV>
   A: <PLNEGSC+VerbRoot+NEGFV>

The system also has question on making the noun singular given the plural (in a short sentence) and to turn a negated verb into the positive, i.e.:

5. Q: <*prefixPL*+stem> <*SGSC*+VerbRoot+FV>
   A: <prefixSG+stem> <SGSC+VerbRoot+FV>

6. Q: <PLNEGSC+VerbRoot+NEGFV>
   A: <PLSC+VerbRoot+FV>

Thanks to the modularized approach of the grammar to facilitate flexible question generation, one also could integrate in a template, say, pluralisation of the sentence combined with positive/negative subject concord; i.e., question/answer pair templates:

Q: `<prefixSG+stem>` `<SGSC+VerbRoot+FV>` `<prefixSG+stem>`
A: `<prefixPL+stem>` `<PLNEGSC+VerbRoot+NEGFV>` `<prefixPL+stem>`

Such a question sentence may be, e.g., *umfowethu uwasha inkomishi* '(my) brother washes the cup' and the requested negative plural is then *abafowethu abawashi izinkomishi* '(my) brothers do not wash the cups'.

In order to separate the core engine from any interface, the system provides the POS tags and noun class tags with the question and computed model answer. This allows further flexibility in display, and, moreover, answer hints. For instance, it could be used to reveal the noun class of the noun to help the learner in figuring out what the plural/singular prefix or subject concord is for that noun, or use aforementioned getVerbGrammar("SC") to provide a hint listing all the possible subject concords.

4. Implementation and basic evaluation

The current system can generate 39501 unique question sentences of two or three words and compute their answers, and scrabble 60 general common conversational sentences. The source code is available at http://www.meteck.org/sw/callCodeZU.zip.

To test the accuracy of the system's output of generating the controlled sentences, we used an oracle, i.e., an isiZulu speaker, to check the appropriate aspects of the text that was generated (pluralization and conjugation have been evaluated [2,15]). A linguist was consulted in the detailed analysis of the results. The meaningfulness of the sentences and the grammatical correctness were the two proxies to determine whether a generated sentence was valid. Accuracy testing was carried out by generating 30 sentences (15 singular, 15 plural) covering each type of template and evaluating its outcome, from which the percentage was calculated, weighing each sentence equally. A sentence received a point only if it was completely free of errors of the particular category being assessed, any semantic or grammatical errors would be grounds for inadequacy. There was space for comments on each sentence.

The raw results were 100% semantically meaningful and 96% grammatically correct for two-words sentences, and (at a first pass) 63% semantically meaningful and 58% grammatically correct for three-word sentences. The single 'error' of a two-word sentence was due to omission of ticking the box of grammatical correctness (as it was checked as semantically correct). The primary reasons for the lower accuracy in the three-word sentences were the words in the corpus and the ported pluraliser and conjugator. For instance, *ushukela* 'sugar' does not have a plural and *-enza* 'do' requires additional phonological condition, which are regular exceptions that are correctly handled in [2, 15]. Others are due to the limited rules in the pluraliser; e.g., NC2 normally takes *aba-*, except when the noun refers to groups of some ethnicities or culture (then the plural prefix is *abe-*), which was not covered in [2]. These issues affected 5 sentences. Debatable words from the corpus in the test sentences are, e.g., -*sheka*, of which it is unclear whether it exists in its own right (meaning: defecate, to be scared, or to commit something) or is an acceptable (or not) colloquial contraction of *shiyeka* 'stay behind', and whether *udadewenu* 'your sister' should be spelled as such or as *udade wenu*. These issues count for 7 cases, which a CALL system is not expected to resolve, but is for linguists and speakers of isiZulu to decide upon. Thus, the CNL templates function exactly as intended, the underlying algorithms perform mostly well, and the word chaining process also works well.

5. Discussion

The NV and NVN template structures with pluralization and negation may look simple from the perspective of isolating languages. For instance, in English, negation amounts to simply 'does not' or 'do not' regardless who or what the subject is and regardless the morphology of the verb. For verb negation in isiZulu, there are 12 singular NCs + 9 plural NC combinations with singulars + 6 personal pronouns = 27 negative SCs to consider and then to remember a set of phonological conditioning rules. Put differently, the range of templates may seem small, but the variability of what can possibly be slotted in is much higher. This thus also entails that canned questions and answers are not feasible even with the current 6 types and 12 templates (aside from shortcomings of not being able to extend it easily).

Overall, the system is a step in the direction of providing many more exercises to the thousands of isiZulu language learners. It already solves the resource issues of limited question sets, of time to mark (instantly cf. weeks), and script loss. We have conducted preliminary experiments with assigning difficulty levels to the exercises—which is integrated in the system presented here—that aims to contribute to assessing the learner's level and progress.

6. Conclusions

Computer-assisted language learning exercises for isiZulu were designed based on novel templates, a small corpus, and algorithms to compute the answers that adhere to the specified answer templates. The agglutinative nature of isiZulu lends itself well for a modular approach so that new templates easily can be configured from existing components of both template elements and of the algorithm snippets for a particular morphological unit. The system has 100% semantic accuracy for two-word isiZulu sentences, but leaves some room for improvement for three-word sentences.

Exercise extensions include the object concord and past tense, a larger corpus, and more comprehensive testing.

Acknowledgements
We thank Zola Mahlaza and Langa Khumalo for their feedback on the exercises and vocabulary, and a section of the INC. This work is based on the research supported in part by the National Research Foundation of South Africa (Grant Number 93397).

References

[1] S. Braun. From pedagogically relevant corpora to authentic language learning contents. *ReCALL*. **17** (2005), 47-64.

[2] J. Byamugisha, C. M. Keet and L. Khumalo, Pluralizing Nouns in isiZulu and Related Languages. In *Proc. of CICLing'16.* Springer LNCS 9623 (2018), 271-283.
[3] M. Chinkina and D Meurers. Question generation for language learning: from ensuring texts are read to supporting learning. *Proc of BEA-12*, Copenhagen, Denmark, Sept 8, 2017.
[4] L. M. da Costa, F. Bond, and X. He, Syntactic Well-Formedness Diagnosis and Error-Based Coaching in Computer Assisted Language Learning using Machine Translation. In *Proc. of NLPTEA2016*. (2016), 107-116.
[5] P. Diab, M. Matthews, and R. Gokool, Medical students' views on the use of video technology in the teaching of isiZulu communication, language skills and cultural competence. *African Journal of Health Professions Education* **8** (2016), 11.
[6] G. Duman, G. Orhon, and N. Gedik, Research trends in mobile assisted language learning from 2000 to 2012. *ReCALL* **27** (2015), 197-216.
[7] C. Gardent and L. Perez-Beltrachini. Using FB-LTAG Derivation Trees to Generate Transformation-Based Grammar Exercices. *Proc. of TAG+11*, Sep 2012, Paris, France. pp117-125, 2012.
[8] K. Harbusch, G. Itsova, U. Koch and C. Kühner. The Sentence Fairy: a natural-language generation system to support children's essay writing. *Computer Assisted Language Learning*, **21** (2008), 339-352.
[9] D. Hatley, P. Hruschka and I. Pirbhai, Process for system architecture and requirements engineering. Addison Wesley. (2013), 45-47.
[10] M. A. Hearst, Can Natural Language Processing Become Natural Language Coaching? In *Proc of ACL 2015 and IJCNLP 2015 (Volume 1: Long Papers)*. ACL (2015) **1**, 1245-1252.
[11] B. Heine and D. Nurse, *African languages: An introduction.* Cambridge University Press, Cape Town, 2000.
[12] HelloTalk, Free Language Exchange, *Language Partners: 2017.* https://www.hellotalk.com/. Accessed: 2017- 09- 30.
[13] C. Henderson, *Building scalable web sites*. O'Reilly California, Sebastopol, 2006.
[14] C. M. Keet and L. Khumalo, Grammar rules for the isiZulu complex verb. *Southern African Linguistics and Applied Language Studies*, **35**(2) (2017), 183-200.
[15] C. M. Keet and L. Khumalo, Toward a knowledge-to text controlled natural language of isiZulu. *Language Resources and Evaluation* **51** (2017), 131-157.
[16] L. Khumalo, Advances in Developing corpora in African languages. *Kuwala*, **1** (2015), 21-30.
[17] L. Kuang-wu, English teachers' barriers to the use of computer-assisted language learning. *The Internet TESL Journal* **6** (2000), 1-8.
[18] C. Lai and W. Kristonis, The advantages and disadvantages of computer technology in second language acquisition. *National Journal for Publishing and Mentoring Doctoral Student Research*, **3**(1) (2006), 6p.
[19] Language-Learning Software and Online Courses, *Transparent Language: 2017.* https://www. transparent.com/. Accessed: 2017-10-01.
[20] S. Makoni, From misinvention to disinvention of language: Multilingualism and the South African Constitution. *Black linguistics: Language, society, and politics in Africa and the Americas.* (2003), 132-155.
[21] F. Mishan, Authenticating corpora for language learning: a problem and its resolution. *ELT Journal* **58** 3 (2004), 219-227.
[22] B. Ndaba, H. Suleman, C.M. Keet and L. Khumalo. The Effects of a Corpus on isiZulu Spellcheckers based on N-grams. *IST-Africa 2016*. IIMC. May 11-13, 2016, Durban, South Africa.
[23] H. Pashler and G. Baylis, Procedural learning: I. Locus of practice effects in speeded choice tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **17** (1991), 20-32.
[24] L. Perez-Beltrachini, C. Gardent, and G. Kruszewski. Generating Grammar Exercises. *7th Workshop on Innovative Use of NLP for Building Educational Applications, NAACL-HLT*, Jun 2012, Montreal, Canada. pp.147-157, 2012.
[25] M. Prabitha, 2010. Computer assisted language learning: Benefits and barriers. *Journal of Literature, Culture and Media Studies* **2** (2010), 59-71.
[26] A. M. G. Sanz, M. Levy, F. Blin and D. Barr, *WorldCALL: Sustainability and computer-assisted language learning*. Bloomsbury Publishing. London. 2015.
[27] J.M. Sinclair (Ed.). How to use corpora in language teaching. *John Benjamins Publishing* **12** (2004).
[28] S. Spiegler, A. Van Der Spuy and P.A Flach, August. Ukwabelana: An open-source morphological Zulu corpus, *In Proc. of COLING 2010*, ACL, 1020-1028.
[29] N.S Turner, *isiZulu Sokuzwana*. University of KwaZulu-Natal. Durban, South Africa, 2011.
[30] R. Vesselinov and J. Grego, J, *Duolingo effectiveness study*. Technical report, City University of New York, 2012.
[31] Y. Xu, A. Goldie and S. Seneff. Automatic question generation and answer judging: a Q&A game for language learning. *Proc of SLaTE 2009*. Warwickshire, UK, Sept 3-5, 2009.