IsiXhosa weather forecast generation

Introduction

The UK's Met Office has investigated the use of natural language generation (NLG) technologies to streamline the production of weather forecasts. A solution inspired by their work would be of great benefit to South Africa because there is no fast and large scale producer, automated or otherwise, of textual weather summaries for Nguni languages (isiXhosa, isiZulu, siSwati, and isiNdebele).

Imozulu eKapa

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

Figure 1: Illustration of a weather forecast

Problem

There are no non-trivial weather forecast generation system for any Nguni language because templates cannot handle the languages' grammatical features, computational grammar rules are few and far between, combinations of templates and computational grammar rules are only ad hoc, and there aren't sufficient corpora for data-driven approaches.

Grammar-infused templates

· New model of templates:
  · **Detachability**: modular design that facilitates grammar re-use across domains
  · **Scaffolding**: the possibility to encode grammar rules within the underlying templates since the existing grammar engines in some languages may be limited if they even exist
· Created isiXhosa weather forecast grammar-infused templates based on GALiweather (Ramos-Soto, 2018)
· Evaluated grammatically and fluency of generated text with 18 L1 isiXhosa speakers
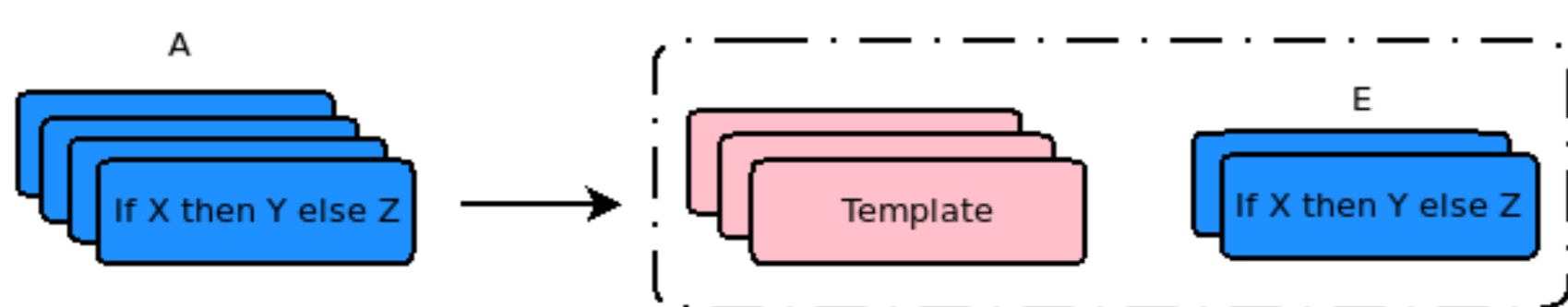

Figure 2: Model of grammar-infused templates highlighting the attachment (A) and embedding (E) relationships

T14: Iqondo eliphantsi lemozulu liphezulu kakhulu kwaye iqondo eliphezulu liphantsi kakhulu xa lithelekiswa nethemprisha elindelekileyo kwelixesha enyakeni, nangona amaqondo azakunyuka esehla ("The temperatures will be very high for the minimums and very low for the maximums compared to the expected for this time of the year, even though they will oscillate")

Figure 3: Example of isiXhosa text generated by the grammar-infused templates

Results and conclusions

· Table 1 shows the number of slot and polymorphic words per template.
· All texts free of morphological agreement and phonological conditioning errors
· 57% of texts perceived as fluent and grammatically correct
· No evidence that the 43% are perceived as unclear or
· grammaticality incorrect, just no consensus

Table 1: Number of slots and polymorphic word in each of the 22 isiXhosa templates

Template ID(s)	Slots	Polymorphic words
1	6	2
2	1	0
3,4	5	2
6,22	3	1
7, 9, 10, 11, 12, 13, 14	3	0
5, 15, 16, 17, 18, 19, 20	2	0
21	2	1
23	4	2

Zola Mahlaza (zmahlaza@cs.uct.ac.za)
Supervisor: A/Prof. C. Maria Keet

Research School
University of Cape Town

HPI

Powered by Hasso-Plattner-Institut