IsiXhosa weather forecast generation

Zola Mahlaza
Supervisor : Assoc. Prof. C. Maria Keet

Computer Science department
Isikolo se-IT

Oct 2020

Outline

1. Natural language generation
2. Problem statement
3. Grammar-infused templates
4. IsiXhosa GALiWeather templates

Natural language generation

- Concerned with the building of systems that can automatically generate text from some provided non-linguistic representation of information (Dale and Reiter 2000)

Input: [[1,SSW,16,20,-,-,0600], [2,SSE,-,-,-,-,NOTIME], [3,VAR,04,08,-,-,2400]]

Output: SSW 16-20 GRADUALLY BACKING SSE THEN FALLING VARIABLE 4-8 BY LATE EVENING

(Example taken from Sumtime project (`https://www.abdn.ac.uk/ncs/departments/computing-science/sumtime-317.php`))

Surface realisation methods

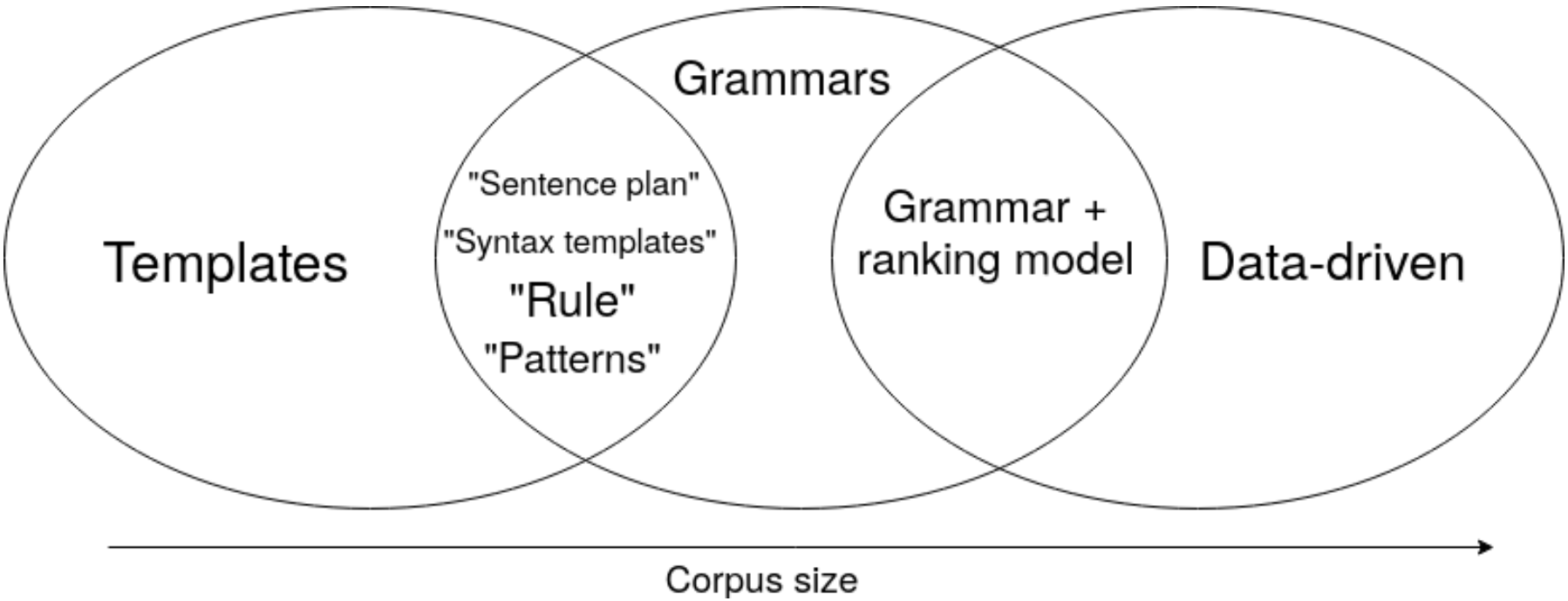

Figure: Current approaches for realisation.

Problem statement

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.

Solution (1/3): grammar-infused templates

- Grammar-infused templates (Mahlaza and Keet 2019)
  - 'Internal/external' pairing
  - Scaffold via grammar rules
  - Re-use grammar rules

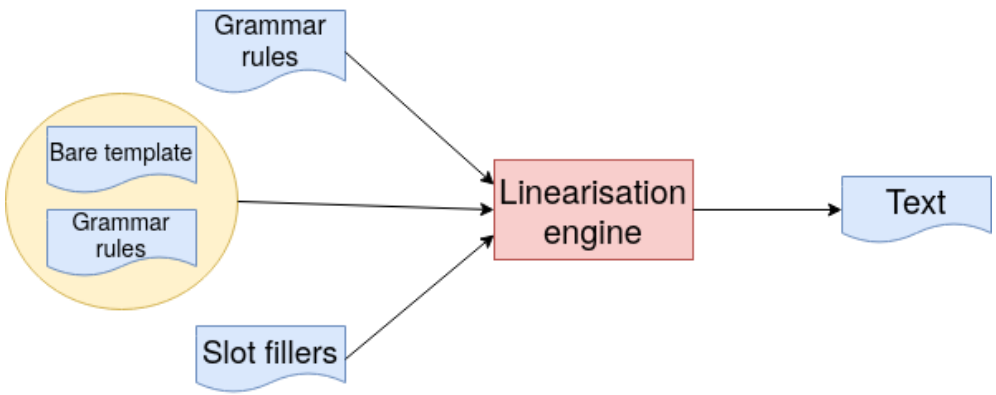

Figure: Text generation with grammar-infused templates

Solution (2/3): grammar-infused templates

- Embedding: computational grammar rule subset ceases to exist on deletion of templates.
- Attachment: computational grammar rule subset loosely coupled
  - Compulsory: every template must use rules from that subset.
  - Partial: not all templates must use rules from that subset.

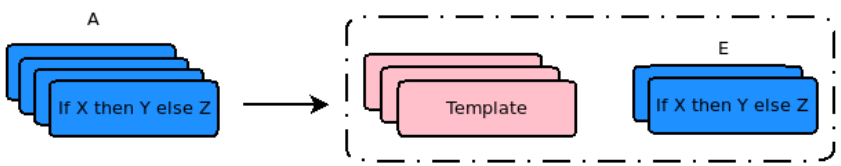

Figure: Model of grammar-infused templates (Mahlaza and Keet 2019)

Solution (3/3): grammar-infused templates

- Model of morphologically enhanced bare templates
- Linearisation algorithm
- IsiXhosa templates for weather forecast generation based on GALiWeather (Ramos-Soto et al. 2015)

Weather forecast evaluation

- Amaqondo obushushu azakuba ephantsikakhulu ngelixesha enyakeni,nangonaamaqondo awazokunyuka esehla ekuhambenikwexesha ("The temperatures will be verylow for this period of the year, even thoughthey will not oscillate over time")
- 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 grammatically incorrect, just no consensus.

References I

Robert Dale and Ehud Reiter. "Building Natural Language Generation Systems". In: *Cambridge University Press* (2000).

Zola Mahlaza and C. Maria Keet. "A classification of grammar-infused templates for ontology and model verbalisation". In: *Metadata and Semantic Research - 13th International Conference, MTSR 2019, Rome, Italy, 28-31 October 2019*. 2019.

Alejandro Ramos-Soto, Alberto José Bugarín Diz, Senén Barro, and Juan Taboada. "Linguistic Descriptions for Automatic Generation of Textual Short-Term Weather Forecasts on Real Prediction Data". In: *IEEE Trans. Fuzzy Systems* 23.1 (2015), pp. 44–57.