

Assignment #4

CS 486 - Fall 2016

Hanumanth Kumar Jayakumar - 20527136

Changqi Du - 20508921

Difei Zhang - 20530592

Jichen Zhao -20453860

December 3, 2016

Experimental Plan

- a. We plan to research natural language resource issues. Our goal is to extend our ontology using Lexicon entries, NL Names and Sentence Plans. We would like to explore text generation using NL resources, and the effect of enriching these resources on the generated text. We will be documenting our steps and results throughout this process so that we have sufficient information to make informal evaluations and educated guesses.
- b. Our primary domain ontology is a knowledge base focused on World War 2. It contains information about the different participants and alliances fighting the war. We also model different engagements in various locations, and the types of units taking part in these engagements. The ontology also contains information about the different types of commanders leading different units, and the types of army men taking part in the war.
- c. We do not plan to use any existing ontologies. However we plan to use the sample Harry Potter ontology, and its associated Natural Language resources as a guide while creating our own. This will serve as a benchmark for enriching our own natural language resources. We also plan to use the '*Generating Natural Language Descriptions from OWL Ontologies: the NaturalOWL System*' document to further our understanding of the NaturalOWL System.
- d. We plan to be able to form fluent sentences about different armies, the battles they took part in, and different army units. If possible, these sentences would be interesting as well, but that also depends on the amount of information present in the ontology. This will be achieved by creating lexicon entries for the required nouns, verbs and adjectives. Following this, we will create NL names for certain lexicon entries, and connect them to classes and individuals. Finally, we will create sentence plans, connect them to object properties, and assess the effect of these changes on the generated text.

The above steps will be carried out for each sentence, or topic that we wish to generate fluent text for. During this process, we will constantly compare and iterate on our changes. This will enable us to view the results of each change, as well as find the best option out of multiple choices of generated text.

- e. We will test the effect of adding more sentence plans and other natural language resources on the generated text. Our hypotheses is that an increase in the natural language resources should lead to more fluent and grammatically accurate text. These resources should help form proper sentences with pronouns and be somewhat similar to language used in daily life.

Description of Parts Implemented

We started our experiments with generating text for the Wehrmacht 18th Army individual. We found that using a maximum graph distance of 2 generated text with richer content compared to a graph distance of 1. One of the problems we faced initially was that inferred object properties did not show up in the generated text. In order to solve this problem, we asserted a couple of previously inferred properties. We also had trouble figuring out how to generate text with the correct article before it. We solved this by creating articles in NL names, and connecting these NL names to individuals or classes. Also, we figured that it is best practice to have the articles agree with the noun they are referring to in terms of plurality, although this is hard to test since in most situations we only have singular nouns.

We followed the plan as described above, beginning with creation of lexicons and NL names. After linking each NL name to an individual or class, we tested its effect on text generation. This iterative process was useful in debugging, since we forgot to add the connection to a certain class a couple of times. (Clicking on the class from the NL names connection dialog does not add the connection ! You need to click on the + button !)

Once we added in the required lexicons and NL names for a sentence, we create a sentence plan, and connected the plan to an object property. A few times, this gave us some interesting results. For example, if the gender of an inanimate object, was left at the default Masculine/Feminine setting, the sentence plan would replace the property owner with *he/she* in the generated text. This resulted in some pretty funny sentences which made no sense at all !

During this process of enriching our natural language resources, we found that we could not use the same name for the lexicon entry and the NL name. We also made good use of the 'duplicate' feature, which enabled us to easily replicate similar lexicon entries or NL names. When playing with the different forms of lexicons such as the singular and plural forms, we found that it is often useful to experiment with wildly distinct strings such as "generalZZZ" instead of "generals" to more clearly see that the proper form is used. Initially, we used strings within Sentence plans for certain verbs which had to be modified from the object property. However, we realized that adding those verbs to the lexicon entries, and utilizing the system provided to BEVLE would make our NL resources more extensible. So, we added in more verbs to our Lexicon, to make it easier to understand and reuse.

During the process of forming accurate sentences for Wehrmacht's 18th Army, we found that each successive addition to the NL resources, enriched our generated text. It added more fluency and grammatical correctness to the text. Our ontology was based on World War 2, so we also had to modify the default isASPEN sentence plan, to reflect the correct tense in the text generated.

Once we generated text for the Wehrmacht 18th Army, we followed the same process as above for the other armies. We then moved on to enriching the resources for the different engagement instances since they had more interesting information. We were able to add a number of sentence plans, lexicons, and NL names in this process. We were also able to rearrange our sentence plan from the default order, to form more grammatically accurate sentences. The *Siege of Leningrad* in particular gave us some really interesting information compared to other instances.

Finally, we added natural language resources for the different types of units. This was interesting because we had focused on individuals before, but we were now working with classes. Since constraints on these classes were part of the generated text, we tried to add resources which would portray them in an accurate manner. However, we are still left with some funny sentences, caused by universal and existential constraints. Overall, we managed to implement all the parts of the plan as envisioned, and learned a lot about language and the difficulties involved in NLG.