

## **Team Members/Project Type**

This project will consist of a survey paper written by Daniel Harrison and Carlos Lazo concentrating on an in-depth examination of Computational Morphological Analysis as it applies to Natural Language Processing. We plan to study the methods that have been used to approach this problem in the past and provide an analysis of their validity as well as describe modern approaches and current research in the area.

## **Description**

The problem of Natural Language Processing (NLP) is one that has interested scientists from the beginnings of the Artificial Intelligence (AI) movement and, in fact, since the beginnings of computing itself. A solution to this problem would allow for a massively more intuitive and efficient interface to modern computers. NLP is difficult because words may mean unrelated things in different contexts and languages, targets of pronouns and adjectives may be unclear, etc.

The modern understanding of NLP can be divided into several sub problems: Morphological analysis, Syntactic analysis, Semantic analysis, Discourse integration, and Pragmatic analysis. Morphological analysis delves into the realm of word components, looking into elements such as prefixes, suffixes, and indicators of plurality. Meanwhile, relationships between words in a linear language sequence and their respective transformation into structures are explored in Syntactic analysis. Once word sequences are properly formulated, Semantic analysis is performed to assign them correct, distinct meanings. Meanings of word sequences have the potential of being dependent upon preceding sentences and may affect future linguistic structures - Discourse integration helps to ensure that the overall discourse of the sentence is logical. Finally, Pragmatic analysis aids in the ensuring coherent reinterpretation of linguistic structures.

Our topic, linguistic morphological analysis, envelops many different areas: lexemes and word forms, inflection vs. word formation, paradigms and morphosyntax, allomorphy and morphophonology, and lexical morphology. The three modern principal approaches to morphology are morpheme-based, lexeme-based, and word-based. In morpheme-based morphology, words are parsed into their respective morphemes, the minimum meaningful units of a language. Lexeme-based morphology generates word forms based on a process which applies inflectional rules to word stems. Finally, Word-based morphology addresses the unwanted, inherent rigidity with the previous methods of analysis by examining generalizations between word forms.

Computational linguistic morphology is an ever-evolving topic in the AI field that will continue to interest researchers in the years to come. We look forward to immersing ourselves in this topic, with the hope of developing a deep understanding of its intricacies and the role it plays in NLP.

## **Resources**

Useful websites:

<http://www.aai.org/AITopics/html/natlang.html>  
<http://www.cs.dartmouth.edu/~brd/Teaching/AI/Lectures/Summaries/natlang.html>  
[http://en.wikipedia.org/wiki/Morphology\\_\(linguistics\)](http://en.wikipedia.org/wiki/Morphology_(linguistics))  
<http://company.yandex.ru/articles/iseg-las-vegas.html>  
[http://www.cs.uchicago.edu/files/tr\\_authentic/TR-2005-06.pdf](http://www.cs.uchicago.edu/files/tr_authentic/TR-2005-06.pdf)