Introduction to Linguistics

Seungwoo Schin

NCSOFT

principia12@ncsoft.com

June 14, 2020

Overview

Language

Language is...

- Combination of signifier and signified. [de Saussure and Baskin, 2011]

What is signifier and signified?

Signifier and Signified

In easy, every-day term,

- Signifier(기표) : Surface of the text. What we actually see.
- Signified(기의) : What does the text mean.

Therefore, to know a language is to know two things;

- How is the surface of the language organized?
- And how is it related to abstract meaning?

A Thousand Miles Begins with a Single Step

... And every texts start with an alphabet! In this perspective, language is

- A set of atomic tokens
- And their sequence

But is it all? For instance,

- hello world! : is a correct English sentence, while
- elloh ldrow! : is clearly not. (wrong words)
- world hello! : is also not. (right words, wrong structure)

Therefore, we can conclude that there are certain **rules** that works as a constraint for generating *legal* sequences.

But it Begins when the Step is Right I

These constraints are called **Grammar**(more of linguistic term) or **Synax**(more of computational term). These rules consist of

- Tokens to words
- Words to Part-of-Speech
- Part-of-Speech to Speech

From here, we assume that we have a large set of human-made texts, called **corpus(말뭉치)**.

But it Begins when the Step is Right II

Words are elements of vocabulary.

Vocabulary is **a set of token sequences that appears together in the corpus**. It might seem very trivial, but it does have some issues...

- hello, world, presuppose : clearly a word
- hello = hell + o?
- presuppose = pre + suppose?

... than, every word can be reduced to set of alphabet a-z, which is very unlikely.

But it Begins when the Step is Right III

In case of the Korean language, things got worse!

One more issue; Zipf's Law. Top 10% of vocabulary constitue of around 90% of whole corpus. Most words are used only once or twice.

Therefore, defining vocabulary itself is not an easy task; just spliting with space does not make things work!

But it Begins when the Step is Right V

Now suppose we have a vocabulary(somehow). Is it done?

- world hello!
- I like eat.

... Apparently not. There are more rules to be considered.

But it Begins when the Step is Right VI

Through observation, we can conclude that

- Some words appear in similar position.
  - I like (pizza—meat—flower—...)
  - I like (eat—run—...) : not

-

And Miles Make Path!

•

de Saussure, F. and Baskin, W. (2011).
INTRODUCTION, pages 1–37.
Columbia University Press.