# Chapter 1


Ryerson University

# NLP and its application

- NLP is the **set of methods** for making human language accessible to computers.
- NLP  is focused on the design and analysis of computational algorithms and representations for processing natural human language.

- **Application examples** : Automatic machine translation , spam/non-spam email classification,  dialog systems ,extracting information from texts, translating between languages, answering questions, holding a conversation, taking instructions ...

Ryerson University

# Themes and perspectives in NLP :

1. Learning and knowledge
2. Search and learning
3. Relational, compositional and distributional perspectives

Ryerson University

# NLP - Learning and Knowledge Perspective:

Combination of knowledge and Machine Learning in NLP

**Using ML to…**

Train **end-to-end systems** that transmute **raw text** into **output structure** like a **summary**, database, or translation.

Provide compositional Sentence organization → ex: meaning of larger units gradually constructed from the meaning of their smaller constituents.

**Using ML for…**

Transforming **text** into a stack of general-purpose **linguistic structures**:
From **subword units** called morphemes:

1. to **word-level** **parts-of-speech**
2. to **tree-structured representations** of **grammar**
3. to **logic-based representations** of **meaning**.

**Supervised ML systems can…**

make use of features or methods like **stemming, parsing, part-of-speech tagging**

# Search and Learning Perspective

Much of the work of NLP → the design of the model Ψ

- Looking at NLP problems from **mathematical point of view**, with two distinct modules as **Search** and **Learning**.
- **Search module :** This module finds the output **ŷ** that gets the best score with respect to the **input x**. *Because the outputs are usually **discrete** in NLP problems, it relies on **combinatorial** optimization.*
- **Learning module :** This module is responsible for finding the parameters **θ**. *Because the parameters are usually **continuous**, learning algorithms generally rely on **numerical** optimization.*

$$\hat{\boldsymbol{y}} = \underset{\boldsymbol{y} \in \mathcal{Y}(\boldsymbol{x})}{\operatorname{argmax}} \Psi(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta}),$$

| Machine learning | |
|---|---|
| $\Psi(\boldsymbol{x}^{(i)}, y)$ | the score for assigning label $y$ to instance $i$ |
| $\boldsymbol{f}(\boldsymbol{x}^{(i)}, y)$ | the feature vector for instance $i$ with label $y$ |
| $\boldsymbol{\theta}$ | a (column) vector of weights |

Ryerson University

# Relational & Compositional Perspective

- The **relational**, **compositional**, and **distributional** perspectives all contribute to our understanding of **linguistic meaning**.
- The **Relational** perspective on **meaning** is the basis for **semantic ontologies** such as WORDNET → the relations that hold between words and other elementary semantic units.
- **Compositional** perspective : The **meaning** of a word is constructed from the constituent parts (journalists → journalist, journal → jour). This principle can be applied to larger units: phrases, sentences, and beyond