Composition

Language is **compositional**: units such as words can combine to create phrases, which can combine by the very same principles to create larger phrases. For example, a **noun phrase** can be created by combining a smaller noun phrase with a **prepositional phrase**, as in *the whiteness of the whale*. The prepositional phrase is created by combining a preposition (in this case, *of*) with another noun phrase (*the whale*). In this way, it is possible to create arbitrarily long phrases, such as,

(1.1) ...huge globular pieces of the whale of the bigness of a human head.[2]

The meaning of such a phrase must be analyzed in accord with the underlying hierarchical structure. In this case, *huge globular pieces of the whale* acts as a single noun phrase, which is conjoined with the prepositional phrase *of the bigness of a human head*. The interpretation would be different if instead, *huge globular pieces* were conjoined with the prepositional phrase *of the whale of the bigness of a human head* — implying a disappointingly small whale. Even though text appears as a sequence, machine learning methods must account for its implicit recursive structure.

- Eisenstein, 2019, Chp 1

Machine learning problem formulation

- The machine learning approach expresses NLP as an optimization problem:

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

where $\mathbf{x} \in X$ is the input

$\mathbf{y} \in Y$ is the output

$\Psi(\cdot) \rightarrow \mathbb{R}$ is a function expressing the learning objective

$f(\cdot)$ is the function, or model, that maps $\mathbf{x}$ to $\mathbf{y}$

$\boldsymbol{\theta}$ parameterizes $f(\cdot)$