



Department of Computer Science  
UNIVERSITY OF COLORADO **BOULDER**



## Distributional Semantics

Advanced Machine Learning for NLP

Jordan Boyd-Graber

SLIDES ADAPTED FROM YOAV GOLDBERG AND OMER LEVY

## Word Representation

---

- Last time, we saw how valuable hidden layers were for representation (much more language today)
- How can we use it for words?

## Word Representation

---

- Last time, we saw how valuable hidden layers were for representation (much more language today)
- How can we use it for words?
- How similar is “pasta” to “pizza”
- Computers often use one-hot representations
- Or fragile knowledge bases

## Word Representation

---

- Last time, we saw how valuable hidden layers were for representation (much more language today)
- How can we use it for words?
- How similar is “pasta” to “pizza”
- Computers often use one-hot representations
- Or fragile knowledge bases
- Distributional Hypothesis (Harris, 1954; Firth, 1957)
- Know the word by the company it keeps

## Intuition (from Boroni)

---

Marco saw a furry little wampimuk hiding in the tree

## Intuition (from Boroni)

---

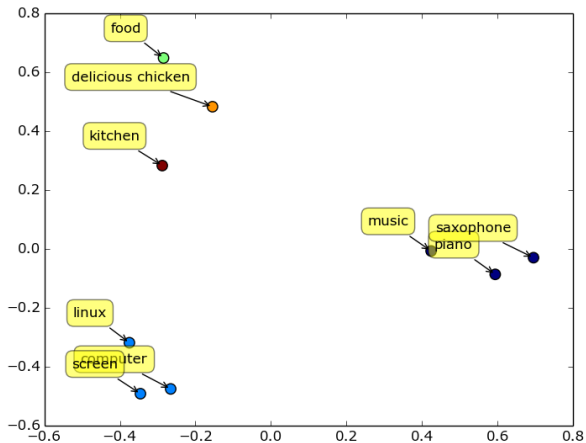
Marco saw a furry little wampimuk hiding in the tree

## Intuition (from Boroni)

---

Marco saw a **furry little** wampimuk **hiding in the tree**

# Representation





## Usefulness

---

- Multimodal
- Multilingual
- Useful downstream feature