



Georgetown
University

Lecture 02: Text Classification

Bag-of-words

- Translation: represent text as a histogram of words
- Strips all syntactic and structural information from the text
- Examples:

demo

Utterance

```
"Mary", "also", "likes", "to", "watch", "football", "games"
```

BOW

```
{"Mary":1, "also":1, "likes":1, "to":1, "watch":1, "football":1, "games":1}
```

Utterance

```
"John", "likes", "to", "watch", "movies", "Mary", "likes", "movies", "too"
```

BOW

```
{"John":1, "likes":2, "to":1, "watch":1, "movies":2, "Mary":1, "too":1}
```