Properties of Collocations in NLP

1 Introduction

Collocations are fixed combinations of words that occur together more often than by chance. They exhibit certain linguistic constraints, including **non-compositionality**, **non-substitutability**, and **non-modifiability**. Below, we explain each property with examples.

2 Non-Compositionality

A collocation is **non-compositional** if its meaning **cannot be inferred** from the individual words.

Examples:

- **Kick the bucket** → (Means "to die")
  The meaning is not derived from "kick" + "bucket."

- **Red tape** → (Means "bureaucratic obstacles")
  The meaning is idiomatic rather than literal.

3 Non-Substitutability

A collocation is **non-substitutable** if **one word cannot be replaced** by a synonym without changing its meaning.

Examples:

- **Make a decision** (Correct)
  Do a decision (Incorrect)
  "Make" cannot be replaced with "do" in this phrase.

- **Heavy rain** (Correct)
  Strong rain (Incorrect)
  "Strong" is a synonym of "heavy," but "strong rain" is not a valid collocation.

4 Non-Modifiability

A collocation is **non-modifiable** if **it cannot be freely changed** by inserting additional words or modifying its grammatical structure.

Examples:

- **Kick the bucket** (Correct)
  Kick the metal bucket (Incorrect)
  Adding "metal" disrupts the idiomatic meaning.

- **Break the news** (Correct)
  Break the shocking news (Incorrect)
  Although "shocking" makes sense, this phrase is rarely modified this way.

5 Summary Table

Property	Meaning	Example
Non-Compositionality	Meaning **cannot be inferred** from individual words	"Kick the bucket" $\neq$ "Kick" + "Bucket"
Non-Substitutability	Words **cannot be replaced** with synonyms	"Make a decision" $\neq$ "Do a decision"
Non-Modifiability	Structure **cannot be changed**	"Kick the bucket" $\neq$ "Kick the metal bucket"

Table 1: Linguistic Properties of Collocations