Understanding Morphology: Add/Delete Table

In [10]:
```
from tabulate import tabulate
```

In [11]:
```
root_words = ['play', 'jump', 'work', 'talk']

modifications = {
    '-ing': 'present continuous',
    '-ed': 'past'
}
```

In [12]:
```
add_results = []
delete_results = []
```

In [13]:
```
for word in root_words:
    for suffix, tense in modifications.items():
        modified_word = word + suffix.lstrip('-')
        add_results.append([word, suffix, modified_word, tense])
```

In [14]:
```
for word, suffix, modified_word, tense in add_results:
    if modified_word.endswith(suffix.lstrip('-')):
        original_word = modified_word[:-len(suffix.lstrip('-'))]
        delete_results.append([modified_word, suffix, original_word, tense])
```

In [15]:
```
print('\nAdd Operation Results:')
headers_add = ["Root Word", "Modification", "Modified Word", "Tense"]
print(tabulate(add_results, headers=headers_add, tablefmt="grid"))
```

```
Add Operation Results:
+-------------+----------------+-----------------+--------------------+
| Root Word   | Modification   | Modified Word   | Tense              |
+=============+================+=================+====================+
| play        | -ing           | playing         | present continuous |
+-------------+----------------+-----------------+--------------------+
| play        | -ed            | played          | past               |
+-------------+----------------+-----------------+--------------------+
| jump        | -ing           | jumping         | present continuous |
+-------------+----------------+-----------------+--------------------+
| jump        | -ed            | jumped          | past               |
+-------------+----------------+-----------------+--------------------+
| work        | -ing           | working         | present continuous |
+-------------+----------------+-----------------+--------------------+
| work        | -ed            | worked          | past               |
+-------------+----------------+-----------------+--------------------+
| talk        | -ing           | talking         | present continuous |
+-------------+----------------+-----------------+--------------------+
| talk        | -ed            | talked          | past               |
+-------------+----------------+-----------------+--------------------+
```

In [16]:
```
print("\nDelete Operation Results: ")
headers_delete = ["Modified Word", "Removed Suffix", "Root Word", "Tense"]
print(tabulate(delete_results, headers=headers_delete, tablefmt="grid"))
```

```
Delete Operation Results:
+-----------------+------------------+-------------+--------------------+
| Modified Word   | Removed Suffix   | Root Word   | Tense              |
+=================+==================+=============+====================+
| playing         | -ing             | play        | present continuous |
+-----------------+------------------+-------------+--------------------+
| played          | -ed              | play        | past               |
+-----------------+------------------+-------------+--------------------+
| jumping         | -ing             | jump        | present continuous |
+-----------------+------------------+-------------+--------------------+
| jumped          | -ed              | jump        | past               |
+-----------------+------------------+-------------+--------------------+
| working         | -ing             | work        | present continuous |
+-----------------+------------------+-------------+--------------------+
| worked          | -ed              | work        | past               |
+-----------------+------------------+-------------+--------------------+
| talking         | -ing             | talk        | present continuous |
+-----------------+------------------+-------------+--------------------+
| talked          | -ed              | talk        | past               |
+-----------------+------------------+-------------+--------------------+
```

In [17]:
```python
# Tokenization and BoW
from nltk.tokenize import word_tokenize
from sklearn.feature_extraction.text import CountVectorizer
import nltk
nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\ASUS\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
```

Out[17]: True

In [18]:
```python
text = "How much wood would a woodchuck chuck could chuck wood,if a woodchuck could chuck wood

tokens = word_tokenize(text)
print("Tokens:", tokens)

vectorizer = CountVectorizer()
X = vectorizer.fit_transform([text])
print("Vocabulary:", vectorizer.get_feature_names_out())
print("BoW Matrix:", X.toarray())
```

```
Tokens: ['How', 'much', 'wood', 'would', 'a', 'woodchuck', 'chuck', 'could', 'chuck', 'wood',
',', 'if', 'a', 'woodchuck', 'could', 'chuck', 'wood']
Vocabulary: ['chuck' 'could' 'how' 'if' 'much' 'wood' 'woodchuck' 'would']
BoW Matrix: [[3 2 1 1 1 3 2 1]]
```