1 Tagging and Tag Sets (10 points)

1.1 When taggers go bad (5 points)

British Left Waffles on Falkland Islands

N V N PREP N N

N N N PREP N N

1.2 Exploring the tag set (5 points)

1.

|  |  |
| --- | --- |
| Number of Tags | Number of Distinct Words |
| 1 | 47328 |
| 2 | 7186 |
| 3 | 1146 |
| 4 | 265 |
| 5 | 87 |
| 6 | 27 |
| 7 | 12 |
| 8 | 1 |
| 9 | 1 |
| 10 | 2 |

2.

2 Viterbi Algorithm (30 Points)

2.1 Emission Probability (10 points)

Frequency Table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | NOUN | VERB | CONJ | PRO |
| ’e | 0.1 | 0.1 | 0.1 | 1.1 |
| ’eg | 0.1 | 0.1 | 1.1 | 0.1 |
| ghaH | 0.1 | 0.1 | 0.1 | 1.1 |
| ja’chuqmeH | 0.1 | 1.1 | 0.1 | 0.1 |
| Legh | 0.1 | 0.1 | 0.1 | 0.1 |
| neH | 0.1 | 1.1 | 0.1 | 0.1 |
| pa’Daq | 1.1 | 0.1 | 0.1 | 0.1 |
| puq | 2.1 | 0.1 | 0.1 | 0.1 |
| qIp | 0.1 | 1.1 | 0.1 | 0.1 |
| rajHom | 1.1 | 0.1 | 0.1 | 0.1 |
| taH | 0.1 | 1.1 | 0.1 | 0.1 |
| tara’ngan | 1.1 | 0.1 | 0.1 | 0.1 |
| yaS | 0.1 | 0.1 | 0.1 | 0.1 |

**Emission Probability:**

|  |  |  |
| --- | --- | --- |
|  | Frequency | Probability |
| ’e | 1.1 |  |
| ’eg | 0.1 |  |
| ghaH | 1.1 |  |
| ja’chuqmeH | 0.1 |  |
| Legh | 0.1 |  |
| neH | 0.1 |  |
| pa’Daq | 0.1 |  |
| puq | 0.1 |  |
| qIp | 0.1 |  |
| rajHom | 0.1 |  |
| taH | 0.1 |  |
| tara’ngan | 0.1 |  |
| yaS | 0.1 |  |

2.2 Start and Transition Probability (5 points)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | NOUN | VERB | CONJ | PRO |
| START | 2.1 | 1.1 | 0.1 | 0.1 |
| N | 0.1 | 3.1 | 1.1 | 2.1 |
| V | 5.1 | 0.1 | 0.1 | 0.1 |
| CONJ | 1.1 | 0.1 | 0.1 | 0.1 |
| PRO | 0.1 | 1.1 | 0.1 | 0.1 |

2.3 Viterbi Decoding (15 points)

1.

2.

3.

4.

5.