Natuurlijke Taalmodellen en Interfaces
Assignment #1

Sharon Gieske & Eszter Fodor
6167667 & ??

April 4, 2013

WAAROM???

1 Introduction

2 Problem Description

3 Approach

4 Results

1. For n = 1, n = 2 and n = 3, submit the list of the 10 most frequent sequences:
   n = 1:

```
5140: ('to',)
4776: ('the',)
4622: ('and',)
4234: ('of',)
3191: ('I',)
2984: ('a',)
2370: ('her',)
2355: ('was',)
2117: ('it',)
2104: ('not',)
```

   n = 2:

```
586: ('to', 'be')
548: ('of', 'the')
424: ('in', 'the')
395: ('I', 'am')
305: ('had', 'been')
285: ('it', 'was')
283: ('I', 'have')
274: ('could', 'not')
264: ('Mr', 'Knightley')
261: ('of', 'her')
```

n = 3:

```
136: ('I', 'do', 'not')
109: ('I', 'am', 'sure')
62: ('a', 'great', 'deal')
58: ('would', 'have', 'been')
55: ('do', 'not', 'know')
51: ('she', 'could', 'not')
50: ('I', 'dare', 'say')
48: ('in', 'the', 'world')
47: ('I', 'assure', 'you')
45: ('could', 'not', 'be')
```

2. For n = 1, n = 2 and n = 3, submit the sum of all frequencies of all sequences for that n.
   sum of n = 1: 159627
   sum of n = 2: 159626
   sum of n = 3: 159625

5 Conclusion