**Semantic Analysis**

*Assignment 3*

*Computational Linguistics*

Cristina Luengo Agulló

Javier Navarro González

École Polytechnique de Louvain

Université Catholique de Louvain

December 09, 2014

**Introduction**

**Lexicon**

After removing the 250 most frequently words from the 5-gram corpus, we obtain the **lexicon** that we are going to use during the assignment. These are the top 20 most frequently words:

|  |  |
| --- | --- |
| Words | Frequency |
| home | 33.482 |
| reason | 33.138 |
| news | 32.772 |
| second | 32.645 |
| difficult | 32.014 |
| job | 31.782 |
| until | 31.475 |
| own | 30.931 |
| pay | 30.586 |
| war | 30.554 |
| big | 30.459 |
| looking | 30.190 |
| white | 30.155 |
| hands | 29.849 |
| show | 29.745 |
| women | 29.717 |
| order | 29.627 |
| beginning | 29.578 |
| different | 29.180 |
| ago | 29.165 |