they gave on the groups of each experiment.

We define PoS coherence as the percent of words that belong to the most common PoS tag in each group. In order to calculate it, all obtained groups are automatically PoS tagged[10]. Then for each group, we count the percent of words that belong to each PoS and identify the most common tag.

2.4.3 Results

Table 2.4 shows the WordNet similarity evaluation. The average similarity score obtained by CLUTO is higher than the score obtained by Word2Vec (0.81-0.96 against 0.67-0.81). This indicates that the distances between the words in the CLUTO groups are shorter and the semantic relations are stronger. Increasing the corpus size improves the results for both CLUTO and Word2Vec. Preprocessing (specifically PoS tagging) improves the obtained results for all of the Word2Vec experiments. The groups obtained using Skip-Gram get lower scores in the evaluation compared with the groups obtained using CBOW.

Table 2.4 Wordnet Similarity

Methodology	Corpus	Similarity
W2V-CBOW	4M (raw)	0.67
W2V-CBOW	4M (lemma)	0.67
W2V-CBOW	4M (pos)	0.72
W2V-CBOW	20M (raw)	0.74
W2V-CBOW	20M (lemma)	0.75
W2V-CBOW	20M (pos)	0.77
W2V-CBOW	40M (raw)	0.77
W2V-CBOW	40M (lemma)	0.78
W2V-CBOW	40M (pos)	0.81
W2V-SG	40M (raw)	0.69
W2V-SG	40M (lemma)	73
W2V-SG	40M (pos)	0.74
CLUTO	4M	0.81
CLUTO	20M	0.92
CLUTO	40M	0.96

with multiple unrelated words"; 4 corresponds to "semantic relation between most of the words in the corpus, without many unrelated words"

[10]We use only the short PoS tag for this evaluation

Table 2.5 shows the results from the expert evaluation of the semantic relations in the groups. The data is similar to the results with WordNet distances. The groups obtained by CLUTO show higher degree of semantic relatedness (2.8-3.4) compared to the groups obtained by Word2Vec (1.6-2.7). The CLUTO groups at 20M and 40M obtain average above 3, meaning that the experts consider all of the groups to be strongly related. For the experiments with Word2Vec, linguistic preprocessing improves the results, especially at biger corpus size (2.5 against 1.8 for 20M and 2.7 against 2 for 40M). The groups obtained using Skip-Gram algorithm are rated lower than the groups obtained using CBOW. The preprocessed corpus obtains better groups, but the difference is smaller than the one observed with CBOW.

Table 2.5 Expert evaluation

Methodology	Corpus	Score
W2V-CBOW	4M (raw)	1.6
W2V-CBOW	4M (lemma)	1.4
W2V-CBOW	4M (pos)	1.8
W2V-CBOW	20M (raw)	1.8
W2V-CBOW	20M (lemma)	2.4
W2V-CBOW	20M (pos)	2.5
W2V-CBOW	40M (raw)	2
W2V-CBOW	40M (lemma)	2.1
W2V-CBOW	40M (pos)	2.7
W2V-SG	40M (raw)	1.7
W2V-SG	40M (lemma)	1.8
W2V-SG	40M (pos)	2
CLUTO	4M	2.8
CLUTO	20M	3.2
CLUTO	40M	3.4

Table 2.6 shows the results for the PoS coherence evaluation. The data shows that the groups obtained from CLUTO are more PoS coherent, compared with the groups obtained by Word2Vec (90-98% against 69-81%). For the corpora of size 20M and above, the groups obtained by CLUTO have almost 100% PoS coherence, meaning that all of the lemmas belong to the same PoS. Both CLUTO and Word2Vec show improved results with the increase of corpus size. The results with Word2Vec indicate that corpus preprocessing largely improves the obtained results (69%-73% against 75%-81%). In fact, for this experiment the corpus preprocessing have bigger impact than the corpus size: a preprocessed corpus with

a size of 4M generates more PoS coherent groups than raw 40M corpus (74-75% against 73%). The experiments with Skip-Gram obtain similar results for raw corpus. For Skip-Gram the preprocessed corpus also obtains better overall results, however lemmatized corpus obtains better results than the PoS tagged corpus.

Table 2.6 PoS coherence

Methodology	Corpus	PoS
W2V-CBOW	4M (raw)	69%
W2V-CBOW	4M (lemma)	74%
W2V-CBOW	4M (pos)	75%
W2V-CBOW	20M (raw)	72%
W2V-CBOW	20M (lemma)	77%
W2V-CBOW	20M (pos)	80%
W2V-CBOW	40M (raw)	73%
W2V-CBOW	40M (lemma)	78%
W2V-CBOW	40M (pos)	81%
W2V-SG	40M (raw)	73%
W2V-SG	40M (lemma)	80 %
W2V-SG	40M (pos)	77%
CLUTO	4M	90%
CLUTO	20M	97%
CLUTO	40M	98%

Overall, all three evaluations identify similar patterns in the obtained clusters: (1) the groups obtained by CLUTO perform better than the groups obtained by Word2Vec; (2) Increasing the corpus size improves the quality of the results for both methodologies. This is true for semantic relatedness as well as for PoS coherence. The tendency to obtain more PoS coherent groups justifies the usage of PoS coherence as evaluation criteria; (3) Linguistic preprocessing improves the quality of the groups obtained by Word2Vec (with both algorithms).

2.5 Conclusions and Future Work

This article compares two methodologies for identifying groups of semantically related words based on Distributional Semantic Models and vector representations. We applied the methodologies to a corpus of English and compared the quality of the obtained groups in terms of semantic relatedness and PoS coherence. We also analyzed the role of different factors, such as corpus size and linguistic preprocessing.