	Words		SAN	SAN Glosses	SAN	Baseline	Best Unsup.	Pagerank
	Mono	Poly	Synsets	Veronis and Ide	Synsets+Glosses		Senseval 2	Mihalcea
File 1 (d00)	**103**	**552**	**0.4595**	**0.4076**	**0.4396**	**0.3651**	unavailable	**0.4394**
File 2 (d01)	**232**	**724**	**0.4686**	**0.4592**	**0.4801**	**0.4211**	unavailable	**0.5446**
File 3 (d02)	**129**	**563**	**0.5578**	**0.4682**	**0.5115**	**0.4303**	unavailable	**0.5428**
Overall	**464**	**1839**	**0.4928**	**0.4472**	**0.4780**	**0.4079**	**0.4510**	**0.5089**

Table 2: Overall and per file accuracy on the Senseval 2 data set.

compare the accuracy of our method against Mihalcea et al.'s on each Senseval 2 file. In this case we included all words, monosemous and polysemous, because we do not have results for Mihalcea et al.'s method on polysemous words only; the reader should keep in mind that these results are less informative than the ones of Figure 3, because they do not exclude monosemous words. There is an overlap between the two

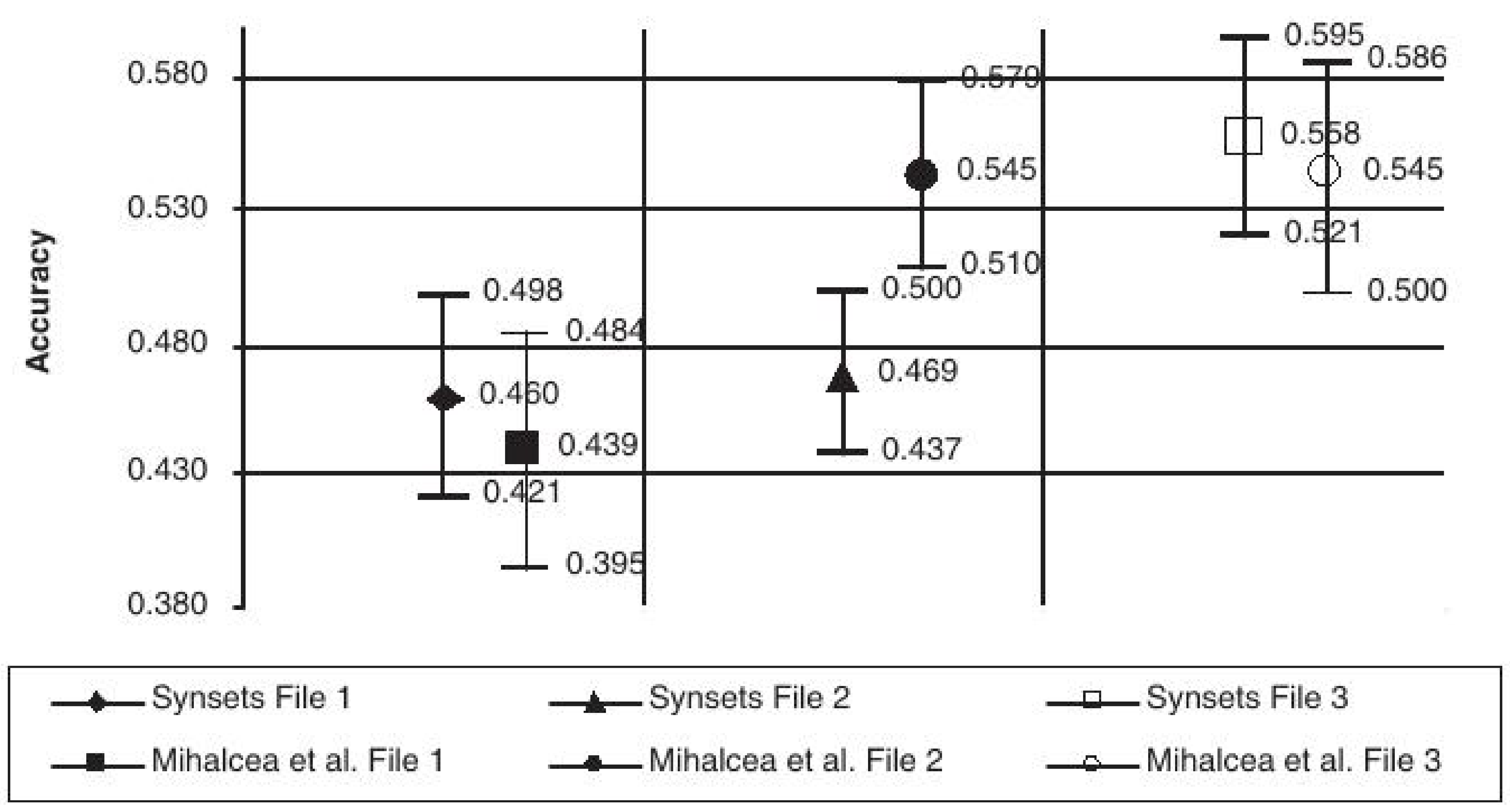

Figure 4: Accuracy on all words and the respective 0.95 confidence intervals.

confidence intervals for 2 out of 3 files, and thus the difference is not always statistically significant.

Regarding the best unsupervised method that participated in Senseval 2, we do not have any further information apart from its overall accuracy, and therefore we rest on our advantage in accuracy reported in Table 2. Finally, we note that to evaluate the significance of our weighting, we also executed experiments without taking it into account in the WSD process. The accuracy in this case drops by almost 1%, and the difference in accuracy between the resulting version of our method and the method of Veronis and Ide is no longer statistically significant, which illustrates the importance of our weighting. We have also conducted experiments in Senseval 3, where similar results with statistically significant differences were obtained: our method achieved an overall accuracy of 46% while Ide and Veronis achieved 39,7%. Space does not allow further discussion of our Senseval 3 experiments.

5.3 Complexity and Actual Computational Cost

Let k be the maximum branching factor (maximum number of edges per node) in a word thesaurus, l the maximum path length, following any type of semantic link, between any two nodes, and n the number of words to be disambiguated. Since we use breadth-first search, the computational complexity of constructing each SAN (network) is $O(n \cdot k^{l+1})$. Furthermore, considering the analysis of constrained spreading activation in [Rocha *et al.*, 2004], the computational complexity of spreading the activation is $O(n^2 \cdot k^{2l+3})$. The same computational complexity figures apply to the method of Veronis and Ide, as well as to the hybrid one, although k and l differ across the three methods. These figures, however, are worst case estimates, and in practice we measured much lower computational cost. In order to make the comparison of these three methods more concrete with respect to their actual computational cost, Table 3 shows the average numbers of nodes, edges, and iterations per network (sentence) for each method. Moreover, the average CPU time per network is shown (in seconds), which includes both network construction and activation spreading. The average time for the SAN Synsets method to disambiguate a word was 1.37 seconds. Table 3

	SAN Synsets	SAN Glosses Veronis and Ide	SAN Synsets+Glosses
Nodes/Net	10,643.74	6,575.13	9,406.04
Edges/Net	13,164.84	34,665.53	37,181.64
Pulses/Net	166.93	28.64	119.15
Sec./Net	13.21	3.35	19.71

Table 3: Average actual computational cost.

shows that our method requires less CPU time than the hybrid method, with which there is no statistically significant difference in accuracy; hence, adding glosses to our method clearly has no advantage. The method of Veronis and Ide has lower computational cost, but this comes at the expense of a statistically significant deterioration in performance, as discussed in Section 5.2. Mihalcea et al. provide no comparable measurements, and thus we cannot compare against them; the same applies to the best unsupervised method of Senseval 2.

6 Related Work

The majority of the WSD approaches proposed in the past deal only with nouns, ignoring other parts of speech. Some of those approaches [Yarowsky, 1995; Leacock *et al.*, 1998; Rigau *et al.*, 1997] concentrate on a set of a few pre-selected words, and in many cases perform supervised learning. In contrast, our algorithm requires no training, nor hand-tagged or parallel corpora, and it can disambiguate all the words in a text, sentence by sentence. Though WordNet was used in our experiments, the method can also be applied using LDOCE