

# Detecting Implicit Aspects in Consumer Reviews for Sentiment Analysis

Kim Schouten, Flavius Frasinca, and Franciska de Jong

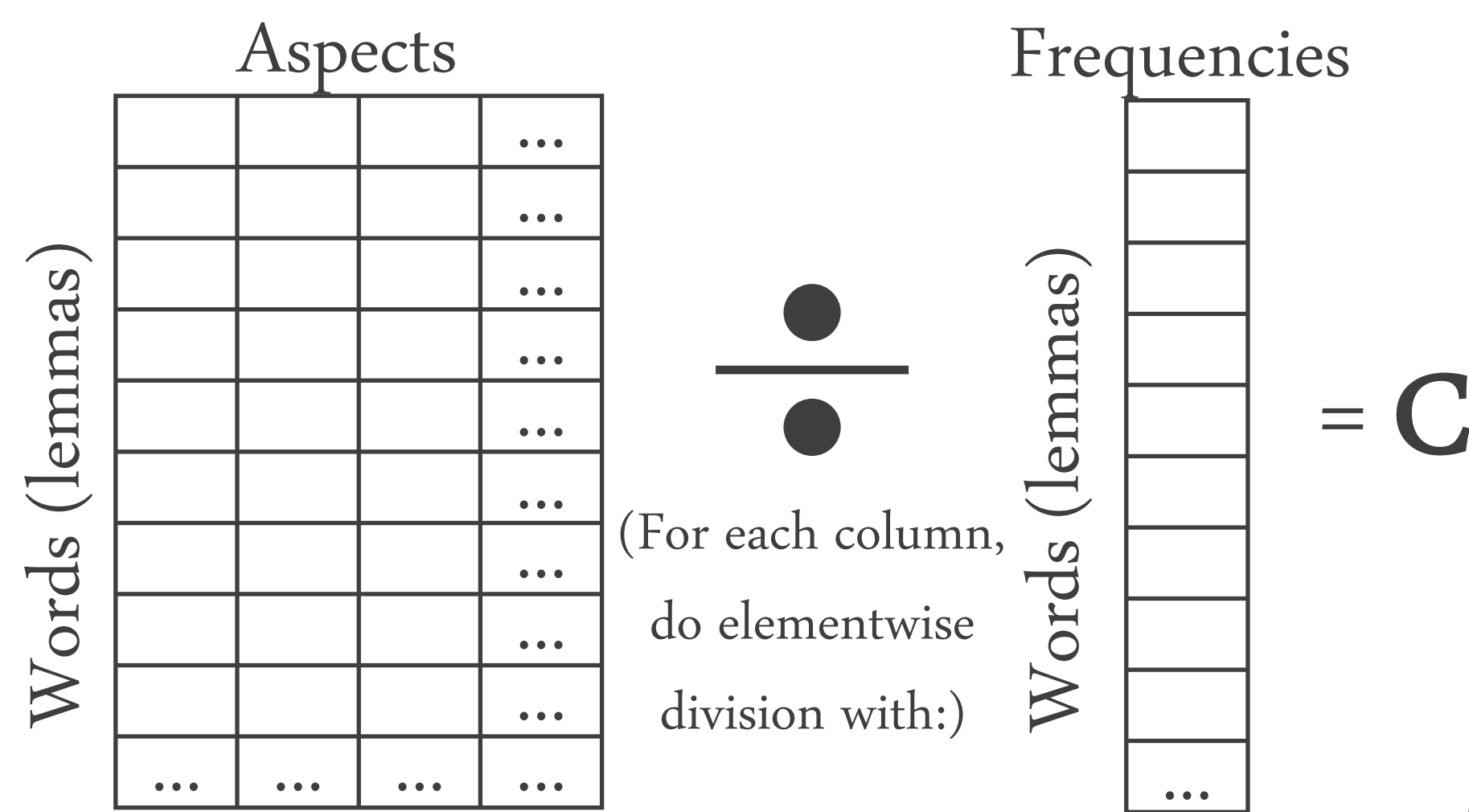
## Word-based co-occurrence approach

### Annotated sentences

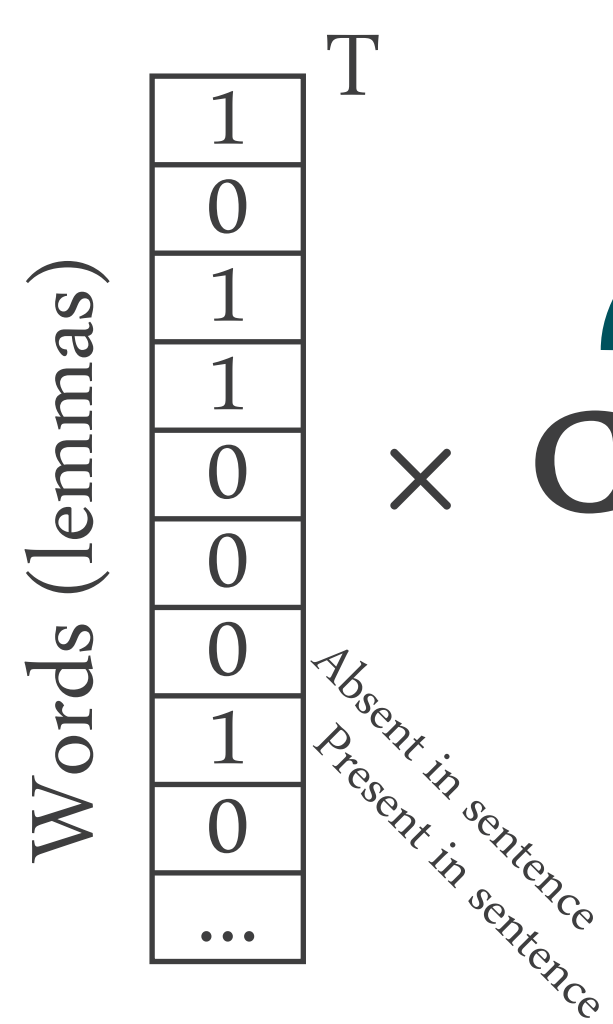


Part-of-Speech filter:  
Only include words that have a specific POS.

### Co-occurrence matrix



### Unseen sentence



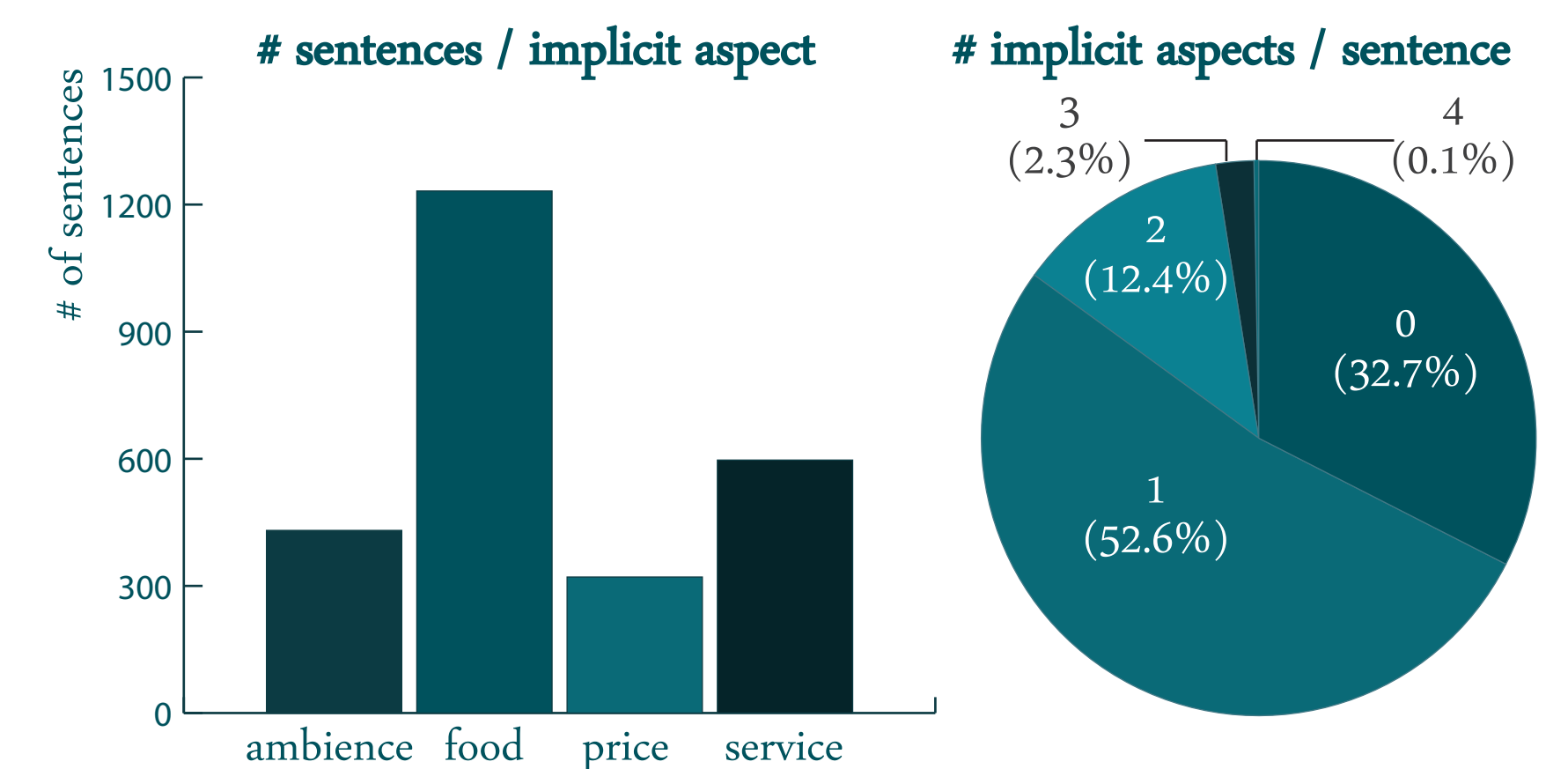
$$T \times C \times \frac{1}{n} = \text{Aspect scores}$$

Number of words in the sentence

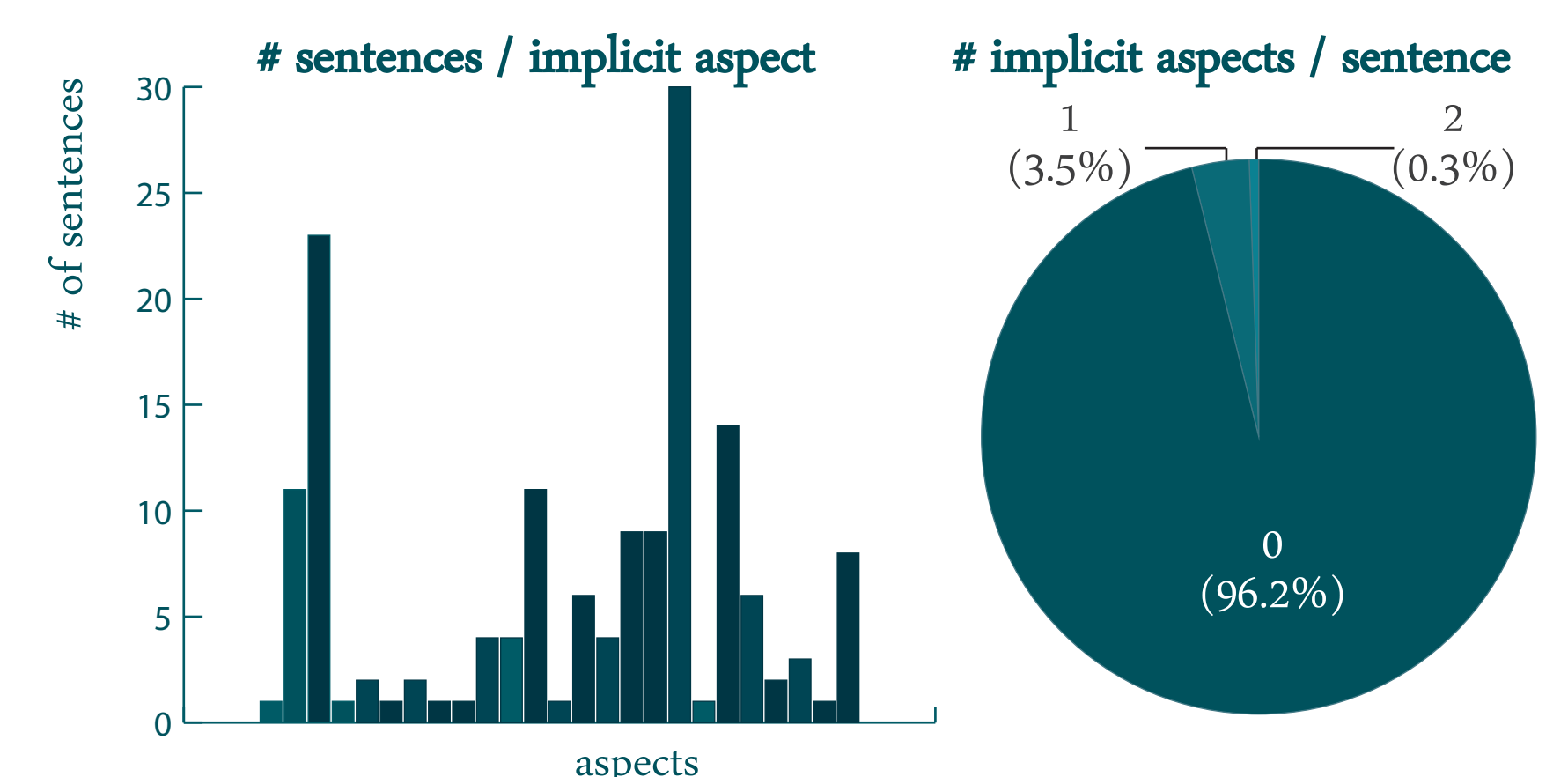
Assign the aspect with the highest score to this sentence as its implicit aspect, if it exceeds a trained threshold value. (threshold is trained on training data only, using linear search)

## Data

### Restaurant reviews (SemEval-2014 training data)

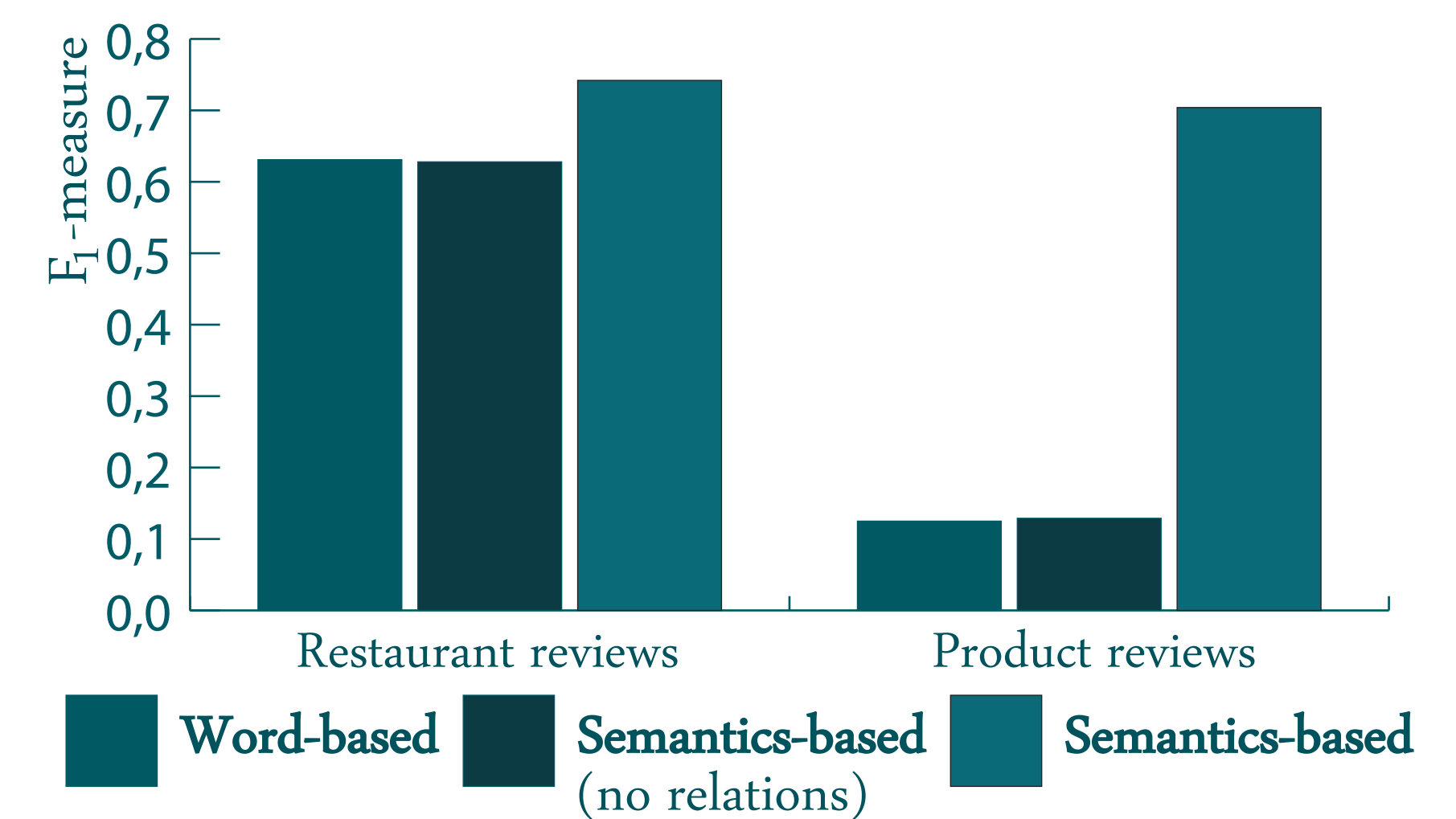


### Electronic product reviews (Hu & Liu, 2004)



## Evaluation

### Results



These results are obtained using 10-fold cross-validation. For each method, the best performing POS filter is used.

### Evaluation & Conclusion

Moving from a traditional word-based to a semantics-based approach is not immediately beneficial. Without including the concept relations, the semantic approach performed roughly the same as the word-based approach. This could be due to errors in the word sense disambiguation. However, when including the concept relations, performance increases significantly.

This is especially the case for the product reviews, a data set which has more different aspects but less training data. We hypothesize that for small data sets, the information added by the concept relations is valuable in detecting implicit aspects. Even for larger data sets, there is still a significant effect.

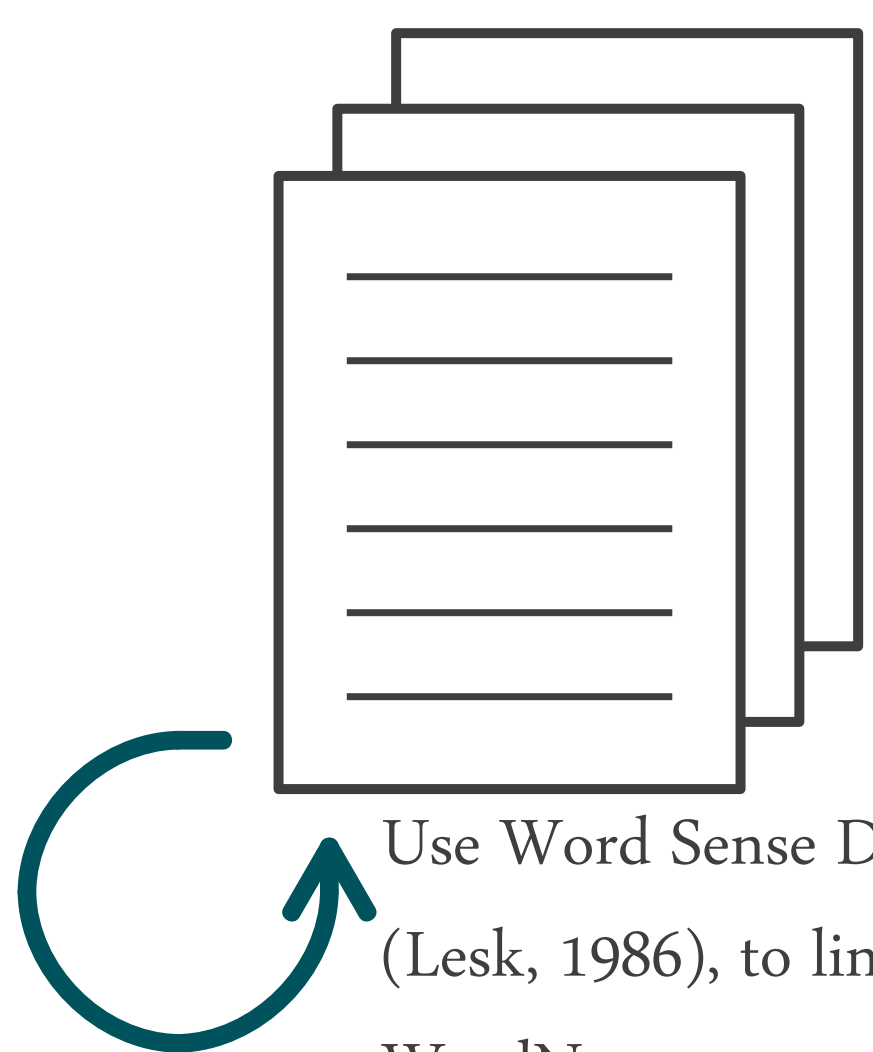
### Future Work

Some of the possibilities for future work are:

1. Adjust the method such that it can detect **multiple implicit aspects per sentence**. This is especially useful for the restaurant data, where a large number of sentences have more than one implicit aspect.
2. **Combine the two training steps**, the genetic optimization for the concept relation weights and the linear search for the threshold value, into one optimization algorithm. A proper training for the POS filter could also be included.
3. **Incorporate domain-specific ontology concepts and relations**.

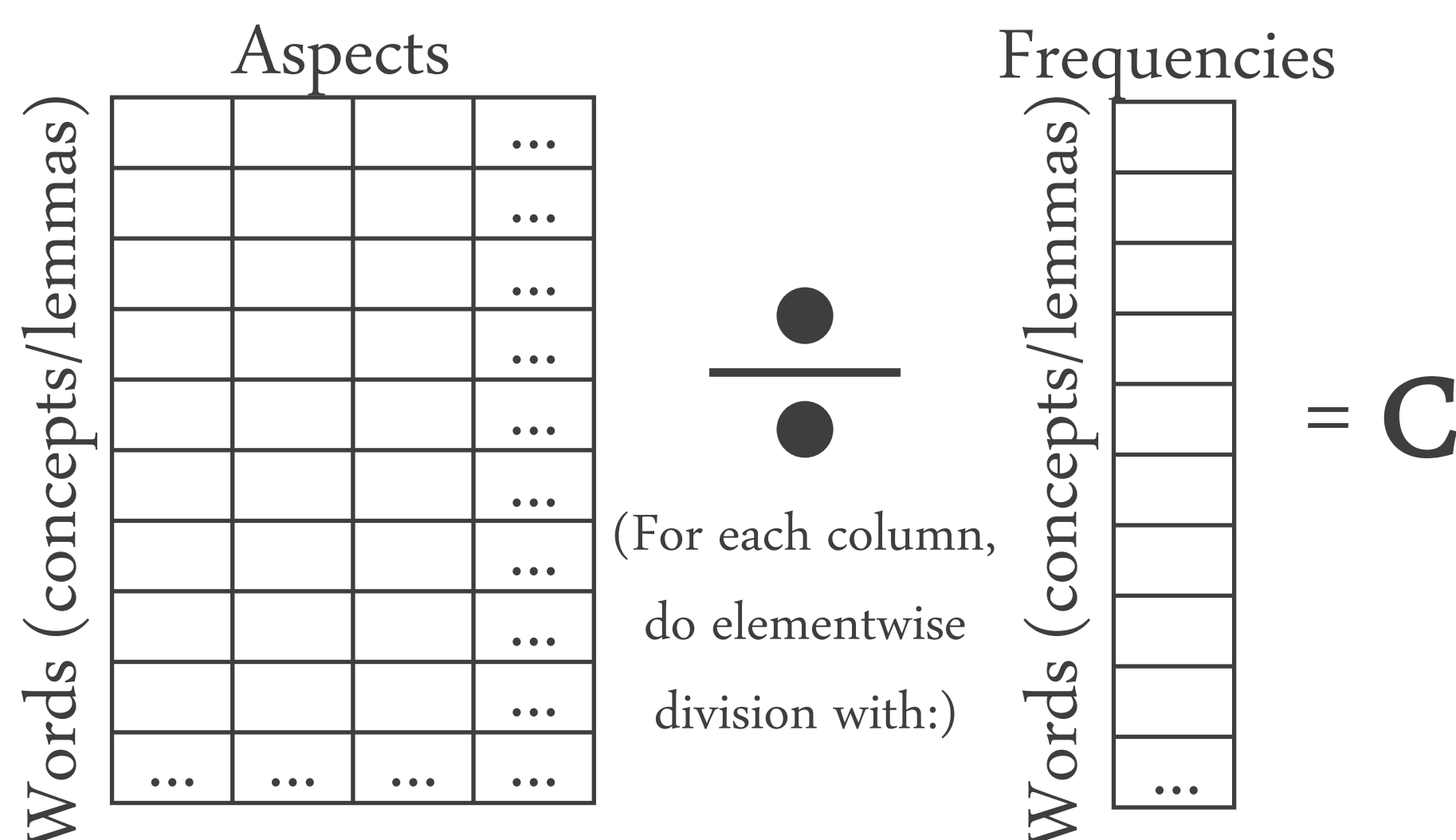
## Semantics-based co-occurrence approach

### Annotated sentences



Part-of-Speech filter:  
Only include words that have a specific POS.

### Co-occurrence matrix



### WordNet concept relations

### Unseen sentence

Words (concepts/lemmas)

Matrix that, for concepts which are directly linked in WordNet, has a value equal to the weight associated with the relation type. It has 1's on the diagonal. (relation weights are between 0 and 1 and are learned using training data only with a genetic optimization algorithm)

= W

$$T \times W \times C \times \frac{1}{n} = \text{Aspect scores}$$

Words (concepts/lemmas)

Assign the aspect with the highest score to this sentence as its implicit aspect, if it exceeds a trained threshold value. (threshold is trained on training data only, using linear search)

### Acknowledgements

The **word-based** method is described in: **Finding Implicit Features in Consumer Reviews for Sentiment Analysis**. Kim Schouten and Flavius Frasinca. In 14th International Conference on Web Engineering (ICWE 2014), volume 8541 of Lecture Notes in Computer Science, pages 130-144. Springer, 2014.

The **semantics-based** method is created together with students Nienke de Boer, Tjian Lam, Marijtte van Leeuwen, and Ruud van Luijk, and is currently submitted for publication.

