# Can Domain Adaptation be Handled as Analogies?

**Núria Bel, Joel Pocostales**
Universitat Pompeu Fabra
Roc Boronat, 138, 08018 Barcelona, Spain
nuria.bel@upf.edu; joel.pocostales@upf.edu

**Abstract**
Aspect identification in user generated texts by supervised text classification might suffer degradation in performance when changing to other domains than the one used for training. For referring to aspects such as quality, price or customer services the vocabulary might differ and affect performance. In this paper, we present an experiment to validate a method to handle domain shifts when there is no available labeled data to retrain. The system is based on the offset method as used for solving word analogy problems in vector semantic models such as word embedding. Despite of the fact that the offset method indeed found relevant analogues in the new domain for the classifier initial selected features, the classifiers did not deliver the expected results. The analysis showed that a number of words were found as analogues for many different initial features. This phenomenon was already described in the literature as 'default words' or 'hubs'. However, our data showed that it cannot be explained in terms of word frequency or distance to the question word, as suggested.

**Keywords:** Aspect identification, domain adaptation, document classification

## 1. Introduction

Machine Learning in general, and classifiers in particular, might suffer degradation in performance when the data to handle belongs to a different domain than the data used for training. Domain adaptation addresses the problem of moving from a source distribution for which we have labelled training data to a target distribution for which we have no labels. Domain adaptation might be crucial for identifying aspects in Aspect Based Sentiment Analysis (ABSA). Note that while for polarity identification, it is likely that a common vocabulary is shared among different domains (e.g. good, bad), for identifying aspects such as quality, design, or support, different domains might exhibit different vocabulary. For instance, quality for a laptop would be described in terms of 'performance': fast, powerful, etc., while for a restaurant would be described in terms of 'taste': delicious, tasty, etc.

In this paper, we present an experiment designed to validate a method to handle domain shifts when there is no available labelled data in a new target domain for an aspect classifier to be retrained. Our approach was to leverage the use of vector space models for semantics such as the one provided by word embeddings (WE, Mikolov et al. 2013). We experimented with using the offset method, as used for solving word analogy problems, to tackle domain adaptation.

In a WE model, we found the examples like in (1), where responses provided by the offset method are in bold.

(1) laptop : shop :: bread : **bakery**

laptop : shop :: beer : **brewery**

laptop : shop :: medicine: **pharmacy**

Thus, we produced lists of analogue words to support domain adaptation of a system for classifying sentences. The task was classifying user-generated texts as according to the aspect (or attribute) of the product (or entity) the user talks about in the text, as described in Aspect Based Sentiment Analysis, ABSA 2014 (Pontiki et al. 2014), Subtask 1, Slot 1: *Aspect Category Detection*.

For our classification experiment, we worked with the following attributes as classes: Design, Price, Quality and Support. While Price was expected not to be very affected by a change of domain, the others would be more affected because differences in vocabularies.

A SVM classifier per aspect, using a bag of words as sentence representations, was trained with user-generated comments on the domain of laptops and tested on the domain of restaurants. In Table 1, some examples and their intended labelling give a hint about the complexity of the task, given the length of the texts. Note that a text can get more than one label.

| | LAPTOPS | RESTAURANTS |
|---|---|---|
| DESIGN | Lightweight and the screen is beautiful! | The music is great, and the lighthearted atmosphere will lifts you spirits. |
| SUPPORT | But no one could tell me when my part would be shipped nor could they tell me where to buy it ON THEIR WEBSITE!!! | We waited for an hour to be seated. |
| QUALITY | The image is great, and the sound is excellent. | The coffee was good even by xxx[1] standards and the food was outstanding. |
| PRICE | It's a steal when considering the specs and performance as well. | Good Food, Great Service, Average Prices. |

Table 1. Selected examples and intended labels.

Despite the fact that the offset method found some relevant words in the new domain, the results of our evaluation experiment showed no improvement with respect to the baseline.

[1] Brands are anonymized

The remainder of the paper is organized as follows. In section 2, we present a review of related research, in section 3, we describe the methodology followed for the experiment; in section 4, the results are presented. In section 5, the results and the error analysis are discussed and, finally, conclusions are presented in section 7.

## 2. Related work

Aspect identification was one of the subtasks of Aspect Based Sentiment Analysis (ABSA) in SEMEVAL 2014 (Pontiki et al. 2014)[2]. The goal was to identify product aspects mentioned in user-generated reviews, for instance if a customer was talking about the quality, price or service of a restaurant. Most teams that participated at SemEval ABSA used SVM classifiers and lexical data as features to represent sentences in different implementations of the bag of words (BoW) approach.

The NRC-Canada system (Kiritchenko et al. 2014), which achieved the best scores (88.57 % F1 and 82.92 % accuracy), used SVMs with features based on various types of n-grams and other lexical information learned from the Yelp dataset. Other systems equipped their SVMs with features that were a linear combination of BoW and WordNet seeds (Castellucci et al. 2014). They used aspect terms extracted using a domain lexicon derived from WordNet and a set of classification features created with the help of deep linguistic processing techniques (Pekar et al., 2014), or they only used BoW features (Nandan et al., 2014). Similarly, Brun et al. (2014) used BoW features and information provided by a syntactic parser to train a logistic regression model that assigned to each sentence the probabilities of belonging to each category. Other teams used the MaxEnt model to build classifiers, where only a BoW was used (Zhang et al., 2014) or they used BoW and *Tf-idf* selected features (Brychcin et al., 2014). Liu and Meng (2014) developed a category classifier with the MaxEnt model with the occurrence counts of unigrams and bigrams words of each sentence as features. Other participating teams only employed WordNet similarities to group the aspect terms into categories by comparing the detected aspect terms either against a term (or a group of terms) representative of the target categories (García Pablos et al. 2014) or against all categories themselves (Bornebusch et al. 2014). Veselovská and Tamchyna (2014) simply looked up the aspects' hyperonyms in WordNet. This approach, however, had many limitations and the systems that used it were ranked in the last positions. And finally, the SNAP system (Schulze et al. 2014) proposed a hybrid approach that combined a component based on similarities between WordNet synsets of aspect terms and categories and a machine learning component, essentially a BoW model that employed multinomial Naive Bayes classifier in a one-vs-all setup. Basically, all the systems made extensive use of lexical data and this creates serious problems when changing the domain.

As for domain adaptation methods, there are a number of different algorithms developed for compensating the degradation in performance. Daumé III (2007) and Blitzer et al. (2006) assumed the availability of some labelled examples in the new domain, and most of the methods proposed after these initial works still require some labelled data of the new domain to retrain, which, in practice, are not available. Daumé III (2007) proposed an approach for supervised adaptation by changing the selected features for ones relevant to the new domain and re-training the classifiers with an augmented list of features.

Our method, explained in next section, proposed to augment the initial list of features by projecting them into the new domain. We formulated the problem as analogy questions.

Word analogy questions have been used to demonstrate that vector space representations consistently encode linguistic regularities (Mikolov et al. 2013, Levy et al., 2014, Linzen, 2016, among others). These linguistic relations are referred as "syntactic", including morphological relations such as verbal base forms and gerund forms, or "semantic" involving world knowledge such as currencies in different countries. Our task was closer to find semantic relations, as we intended to find words expressing specialization of taxonomic relations, for instance finding parts-of or properties-of. To our knowledge, this is the first time that word analogy method is applied to a domain shift problem.

## 3. Methodology

Our proposal worked upon a basic text classification approach that, as we have seen in section 2, used a supervised classifier with features extracted from the corpus represented as a bag of words (BoW). A classifier was trained for every class or aspect, as listed in table 1, and testing was done as one-vs.-all setup. Note that a text can get more than one label.

### 3.1 Reduced BoW feature selection

The BoW representation of texts has been successfully used for document classification. However, for short text classification, this approach delivers very sparse vectors, which are not useful for classification purposes. Different techniques have been devised for vector dimensionality reduction, among these, the ones based on statistical feature selection according to an observed training dataset. In our experiment, we used Adjusted Mutual Information, AMI (Vinh et al. 2009), and chi-squared test to select the words for representing sentences. While AMI, and in general Mutual Information based measures, are known to be useful to identify relevant features, they are biased towards infrequent words. To compensate this bias, we combined it with chi-squared selected ones. Thus, our system first ranks the best candidates in two separated lists, each using a different measure. Then, the two lists are joined into a new one by summing the AMI and chi-squared scores[3]. For instance, if a word is ranked 3rd by AMI and 5th by chi-squared, in the joined list it will be the 8th. A single BoW of 600 features was used for all the aspect classifiers.

For the classifiers, we trained SMO classifiers (as implemented by Weka, Hall et al., 2009). Texts were processed as follows. First, they were cleaned eliminating urls, hashtags, and rare characters. Second, texts were

[2] ABSA 2016 presented an Out-of-Domain track only for French, but no participants registered for it.

[3] In case of tie, results are ordered alphabetically.

tokenized and lemmatized using Freeling 4.0 (Padró & Stanilovsky, 2012). Stop words were eliminated before assessing the combined AMI+chi-squared rank explained before. Note that brand names were also ignored and were not selected for the BoW if recognized. Once the list of selected words is obtained, another module read texts and converted them into 600 dimension binary vectors.

### 3.2 Mapping features to the new domain

The domain adaptation experiment was based on this reduced BoW. For each feature in it, analogous words in the new domain were found by applying the vector space-based offset method in the following way.

(2) laptop : [each feature] :: meal : X

Then, when converting the new domain sentences into a vector, the occurrence of either the initial feature or the found X was considered a positive feature. In this way, no retraining of the classifiers would be necessary, and the classifiers would have to perform well in both domains.

Note that for aspect identification, to retrieve a related word, although not exactly a corresponding analogue word, should be enough as the goal is to take into account words that refer to a particular aspect of a product. It could be different for polarity analysis where it is not the same to observe 'good' than 'bad'. But for aspect identification both, even if antonymous, refer to quality, for instance.

For computing the offset, we used the 3COSMUL method as proposed by Levy et al. (2014). 3COSMUL was demonstrated to better balance the different aspects of similarity to prevent that similarity aspects in different scales can be more predominant in the calculation. The list of analogues proposed by 3COSMUL, which comes from all the corpus vocabulary, was filtered by discarding stop-words and forms not found in a spelling dictionary. Therefore, the list of features used for the out-of-domain classification experiment included the initial ones and the features that were ranked first by 3COSMUL that were actual words (preventing, for instance, forms such as *tablespoonful*) and were not prepositions, pronouns, etc.

To create the vector space model to extract WE, a ten window `word2vec` Skip-Gram (Mikolov et al., 2013) with negative sampling model was trained with the following corpora: a Wikipedia dump[4] and the training initial domain datasets totalling 636M words. Other parameters were: algorithm SGNS, 300 dimensions, context window = 10, subsampling $t{=}10^{-4}$, context distribution smoothing = 0.75, and 15 iterations.

### 3.3 Evaluation Datasets

We used the ABSA 2016 (Pontiki et al. 2016) datasets for English on *laptops* and *restaurants*. ABSA 2016 proposed a closed list of aspect or attribute labels for each product or entity. Our aim was using the classifier trained for a particular domain (laptops) to classify texts of another domain (restaurants) and therefore a common set of labels was needed. Moreover, ABSA entities are very fine-grained: the restaurant corpus included six entity labels (i.e. restaurant, food, drinks, ambience, service, location) and the laptop corpus included 22 (i.e. laptop, display, keyboard, mouse, motherboard, cpu, fans_cooling, ports, memory, power_supply optical_drives, battery, graphics, hard_disk, multimedia_devices, hardware, software, OS, warranty, shipping, support, company). Therefore, for the experiments reported here, only the laptop and restaurant entities were considered. As for attributes, we used DESIGN, PRICE, QUALITY and SUPPORT (Vázquez et al. 2014, Bel et al. 2017) as labels for general aspect identification. ABSA entities and attributes were automatically mapped to these labels as follows:

- LAPTOP DESIGN_FEATURES and RESTAURANT STYLE attribute labels (used for all the texts that include a reference about specific features such as size, color, presentation, styling, ambience, etc.) were directly relabeled as a single DESIGN label.
- LAPTOP PRICE and RESTAURANT PRICES, for texts that comment on prices of goods or services, were relabeled as PRICE.
- QUALITY attributes, for texts that refer to the quality, performance or positive and negative characteristics of a product or service that affects user experience were used as our QUALITY label.
- Both LAPTOP SUPPORT, for pre- and after-sales customer support, repair services and staff, and RESTAURANT SERVICE, for opinions focusing on the service in general, staff's attitude and professionalism, etc., were merged into a common SUPPORT label.

As already mentioned, the laptop corpus was used for training (but a small held out dataset for testing) and the restaurant corpus for testing the out-domain scenario. Table 2 shows the distribution of the datasets used for the experiment.

| | Training | Testing | |
|---|---|---|---|
| Aspect | | ID-HO | OD |
| DESIGN | 344 | 105 | 57 |
| QUALITY | 378 | 87 | 238 |
| SUPPORT | 144 | 33 | 144 |
| PRICE | 136 | 25 | 44 |
| NONE | 2601 | 323 | 103 |
| TOTAL | 3603 | 573 | 587 |

Table 2. Size of the datasets in sentences. ID-HO for In-Domain Held Out and OD for Out Domain. NONE label is for texts with other labels, ignored in this experiment, or with no labels

## 4. Results

Table 3 shows the results of the classifiers, in a one-vs-all scenario, for testing with in-domain data as well as out-domain data and out of domain data represented with

[4] Snapshots of 19-03-2016

vectors created with the initial and analogue feature list. Analogues found by the offset method were added to the initial selected feature list when converting sentences into vectors, as explained in section 3.2.

| | In Domain | | | Out Domain | | | Analogues | | |
|---|---|---|---|---|---|---|---|---|---|
| | P | R | F1 | P | R | F1 | P | R | F1 |
| DESIGN | 0.27 | 0.56 | 0.36 | 0.10 | 0.19 | 0.12 | 0.08 | 0.28 | 0.12 |
| QUALITY | 0.30 | 0.67 | 0.42 | 0.49 | 0.22 | 0.30 | 0.39 | 0.20 | 0.26 |
| SUPPORT | 0.19 | 0.42 | 0.26 | 0.68 | 0.51 | 0.58 | 0.46 | 0.60 | 0.52 |
| PRICE | 0.25 | 0.84 | 0.39 | 0.40 | 0.38 | 0.39 | 0.34 | 0.36 | 0.35 |

Table 3: Results of the classification experiment

As expected, the performance of the classifiers was affected by the change of domain. For three of the four categories, the most relevant impact was a noticeable decrease in recall, because of the lexical differences in these two domains. As expected, the category PRICE was not much affected, in general terms, because of the little change of vocabulary, as already mentioned, original features are also used. Surprisingly, the SUPPORT category performed better in the out-domain scenario, but it might be because this category is more clearly expressed in the restaurant data. Compare restaurant examples as in (2) with laptop examples as in (3), for instance (in bold, words in the initial list of selected features).

(2) We **never** had to **wait** more than 5 **minutes**.

The **service ranges** from mediocre to offensive.

(3) It **took** 3 **days** to **make** an appointment at the local **store**.

I could not **believe** they did not **consider** the **battery** as defective so I **went** to the **store** myself and **asked** for a manager.

As for the out-domain dataset using sentence representation that took into account the suggested analogous words, the results did not show the expected improvements in recall, nor a consistent improvement with respect to the out-domain simple test. We discuss these results in the next section.

## 5. Discussion

Despite of the fact that the offset method indeed found relevant words on the new domain, as shown in Table 4, the classification results did not show the expected improvement. We performed an error analysis addressing two questions: to what extent the analogy questions indeed retrieved words related to the new domain and therefore could be informative, and to what extent found analogues were good features for the classifier.

### 5.1 Are selected words analogues?

It has already been said that, for the aspect classification task, we expected to find a method for mapping, in a loose way, specialized words from one domain to another. Basically, we expected that there would be some words pairs like the ones in (1) that would map. The intuition was that the analogy method could find related words that without being real analogues, nevertheless, could be useful for aspect classification. There would be lexical relations such as 'part-of'' that would map laptop components to meal components, or cases like *driver-baguette*, *keyboard-accompaniment*, and others shown in table 4. For other words that were not particularly related to any of the domains, we expected that the method would select near synonyms, like the examples in table 5, which are also actual mappings in our experiment.

| Laptop | Meal |
|---|---|
| hour | mealtime |
| delivery | take-out |
| quality | palatability |
| problem | undernourishment |
| computer | food |
| outlet | grab-and-go |
| nightmare | ravenous |
| absolutely | scrumptious |
| store | grocery |
| ergonomics | dietetics |
| house | tavern |
| keyboard | accompaniment |
| premium | all-you-can-eat |
| book | cookbook |
| shop | bakery |
| player | appetizer |
| driver | baguette |
| headphone | grill |
| wire | sirloin |
| chat | mealtime |
| office | restaurant |
| bar | restaurant |

Table 4: Examples of analogues proposed for domain words

| Laptop | Meal |
|---|---|
| travel | trip |
| trip | journey |
| dislike | disgust |
| area | region |
| probably | likely |
| start | begin |
| begin | start |
| build | construct |

Table 5: Example of analogues proposed for general domain words

| Laptop | Meal |
|---|---|
| surf | barbecue |
| company | beverages |
| email | dinner |
| adaptor | fatty |
| box | oatcake |
| edit | paella |
| print | ragout |

Table 6: Example of analogues in the new domain but with a non-obvious relation to the feature

After manual evaluation, we observed that indeed most of the selected analogues belong to the new domain, although sometimes the relation between the feature and the analogue was not obvious, as the selected examples in table 6 shows.

A more quantitative evaluation of the results to assess the mapping to the new domain was not possible because of the impact of the hubness problem that we describe in the next section.

### 5.2 Are analogues good classification features?

In order to evaluate how good the selected words were for classifying the restaurant texts, we run the feature selection method, described in section 3, with the restaurant corpus and we compared the resulting list of 423 selected features with the list of features resulting of the laptop corpus and with the list of proposed analogues.

Laptop domain feature list and the restaurant one shared 201 words, what supports our decision of keeping features selected for the laptop domain also when classifying the new domain.

The list of proposed analogues and the list of selected features for the restaurant corpus shared 48 types, that is, 48 unique words out of the list of 423 features. Some words were suggested as analogues for many different features. Figure 1 plots the number of repetitions for the 58 unique words that were suggested. For instance, *food* was suggested as analogue for 46 different features, *high-carbohydrate* for 40 and *eat* for 81, while *dinner*, *grill*, or *sirloin* were selected for two each.

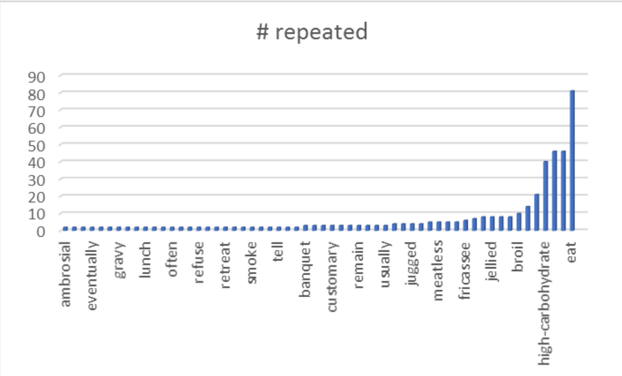


Figure 1: Number of repetitions of the 58 analogues that were suggested for a total of 423 features

For our classifiers, these repetitions caused the classifier vectors to be totally inconsistent and obviously impacted their accuracy rate.

This repetition phenomenon has already been discussed in the literature: Levy et al. (2014) call them "default words" and Dinu et al. (2015), “hubs”. Dinu et al. (2015) found it in the task of English into Italian bilingual lexicon induction. When inducing a function from one language vector space to another language vector space, it was already found that neighborhoods surrounding mapped vectors contained many items that were, in their terms, “universal” neighbors which were called “hubs”. Dinu et al. noted that the hubness problem was exacerbated when there was a mapping from one original space to a target space.

Levy et al. (2014) found it in the analogy task and defined the problem as “one central representative word is provided as an answer to many questions of the same type” and, in this work, it was observed both for explicit and embedding word representations, accounting for the 39% of the errors[5]. In our case, default words represented 43% of the errors. Seven default words (*broil*, *multi-course*, *hearty*, *high-carbohydrate*, *food*, *non-halal* and *eat*) were repeated more than ten times. However, in our results, in addition to clear central representative words (like *food* and *eat*) there were also words that can hardly be considered representative, even although there was a certain degree of similarity for the features that got the same analogue. In table 7, some examples of these cases are shown. Note that for *hearty* most of the analogues seem to be related to ‘emotional’ concepts, for *appetizer* to images, although for *broil* a common characteristic is not obvious.

| hearty | broil | appetizer |
|---|---|---|
| little | battery | graphic |
| charm | processor | card |
| reputation | windows | pro |
| fan | install | video |
| well | add | download |
| nicely | charger | player |
| satisfy | pc | beats |
| hope | heat | bonus |
| disappointment | resolve | |
| definitely | reinstall | |
| astonish | | |
| appreciate | | |
| surprise | | |
| enjoy | | |
| truly | | |
| disappoint | | |
| impress | | |

Table 7: Examples of features that got the same analogues

Linzen (2016), who revised the consistent encoding of lexical semantic relations in vector semantic spaces, found that for most of the cases, the offset method, or more precisely the use of cosine distance for assessing it, tends to retrieve the word which is closest to the query word, in our case *meal*. Furthermore, Schnabel et al. (2015) observed that there is a strong correlation between the frequency of a word and its position in a ranking of nearest neighbors.

However, as table 8 shows, not all the hub words were among the 11 closest words to meal (cf. 'other hubs') and

[5] Levy et al. (2014) define default behavior error when the same incorrect answer is returned for a particular relation 10 or more times.

some of the closest words were never selected (i.e. *breakfast*). As for frequency, although some are very frequent[6] (i.e. *food*) others are rather infrequent words (i.e. *multicourse*). Note that *meal* itself, with a relative frequency of 42.8 per million, is in the same frequency range than *eat*, which is the most suggested analogue. However, it still needs further investigation to find what are the conditions that make other hubs to appear.

| Closest words to *meal* (descending) | RF | # | Other hubs | RF | # |
|---|---|---|---|---|---|
| lunch | 9.60 | 2 | broil | 0.18 | 10 |
| multicourse | 0.01 | 14 | hearty | 0.60 | 21 |
| soup | 7.30 | 8 | high-carbohydrate | 0.03 | 40 |
| food | 132.20 | 46 | | | |
| dinner | 17.80 | 2 | | | |
| eat | 45.30 | 81 | | | |
| breakfast | 11.60 | 0 | | | |
| bread | 13.30 | 0 | | | |
| snack | 4.40 | 0 | | | |
| three-course | 0.06 | 2 | | | |
| non-halal | 0.02 | 46 | | | |

Table 8: List of closest words to 'meal' in descending order. RF stands for relative frequency in percentage per million words, and # for number of repetitions.

## 6. Conclusion

In this paper, we have presented the results and analysis of an experiment that approached domain adaptation as a search for analogues of the selected features of a reduced BoW used to train a SVM classifier. The benefit of such approach would be that the classifier could be used for a new domain without retraining it with a new domain labelled dataset. The results have shown that 3COSMUL (Levy et al., 2014), used in a vector space created with word2vec (Mikolov et al. 2013), found analogues which were relevant for the new domain in a task of aspect identification. However, the phenomenon known as 'default' words or 'hubs', in which some few analogues are selected for many original question words, makes the resulting vectors a bad input for the classifiers. After an error analysis, we found that, contrary to what has been published before, those analogues which are proposed for many features are, although words of the new domain, either very infrequent words, and not generic words (Levy et al., 2014), or words that were not among the closest to the query words (Linzen, 2016). Currently, there is a growing interest in understanding the characteristics of the WE vector space, and how operations like vector offset for finding analogues actually work (Gittens et al. 2017). In future work, we will explore in our data the hints provided by these works to seek a method to improve the list or to filter it such that it allows to create an easy and cheap method for domain adaptation.

## 7. Acknowledgements

This work was supported by the Spanish TUNER project TIN2015-65308-C5-5-R (MINECO/FEDER, UE).

## 8. Bibliographical References

Bel, N.; J. Diz-Pico, M. Marimon, J. Pocostales (2017) Classifying short texts for a Social Media monitoring system, Procesamiento del Lenguaje Natural, Revista nº 59, 57-64.

Blitzer, J.; McDonald, R.; and Pereira, F. 2006. Domain adaptation with structural correspondence learning. In Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP '06). Association for Computational Linguistics, Stroudsburg, PA, USA, 120-128.

Bornebusch, F.; G. Cancino; M. Diepenbeck; R. Drechsler; S. Djomkam; A. Nzeungang Fanseu; M. Jalali; M. Michael; J. Mohsen; M. Nitze; C. Plump; M. Soeken; M. Tchambo; and H. Ziegler. 2014. “itac: Aspect based sentiment analysis using sentiment trees and dictionaries”. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland: ACL and Dublin City University, pages 351–355.

Brun, C.; D. N. Popa; and C. Roux. 2014. “Xrce: Hybrid classification for aspect-based sentiment analysis”. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland: ACL and Dublin City University, August 2014, pages 838–842.

Brychcín, T.; M. Konkol, and J. Steinberger. 2014. “Uwb: Machine learning approach to aspect-based sentiment analysis”. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland: ACL and Dublin City University, August 2014, pages 817–822.

Castellucci, G.; S. Filice; D. Croce; and R. Basili, R. 2014. “Unitor: Aspect based sentiment analysis with structured learning”. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland: ACL and Dublin City University, August 2014, pages 761–767.

Daume III, H. 2007. Frustratingly easy domain adaptation. In Conference of the Association for Computational Linguistics (ACL), Prague, Czech Republic.

[6] Frequency, expressed in percentage per million, was assessed at the English Wikipedia 2014 Corpus of 1.3 billion words using SketchEngine. https://www.sketchengine.co.uk/

Dinu, G.; A. Lazaridou, and M. Baroni, 2015. Improving zero-shot learning by mitigating the hubness problem. In Proceedings of ICLR Workshop Track, San Diego, CA.

García Pablos, A.; M. Cuadros, and G. Rigau. 2014. "V3: Unsupervised generation of domain aspect terms for aspect based sentiment analysis". In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland: ACL and Dublin City University, August 2014, pages 833–837.

Gittens, A., Achlioptas, D., & Mahoney, M. W. (2017). Skip-Gram − Zipf + Uniform = Vector Additivity. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (pp. 69–76)

Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I.H. 2009. The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.

Kiritchenko, S., X. Zhu, C. Cherry, and S. Mohammad. 2014. "Nrc-canada-2014: Detecting aspects and sentiment in customer reviews". In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland: ACL and Dublin City University, August 2014, pages 437–442.

Lazaridou, A., Dinu, G., & Baroni, M. (2015). Hubness and pollution: Delving into crossspace mapping for zero-shot learning. In Proceedings of the 2015 Conference Association for Computational Linguistics, (ACL) Stroudsburg, PA, USA,, pp. 270-280.

Levy, O., Goldberg, Y. and Ramat-Gan, I. 2014. Linguistic regularities in sparse and explicit word representations. CoNLL-2014.

Linzen. T. (2016). Issues in evaluating semantic spaces using word analogies. In Proceedings of the First Workshop on Evaluating Vector Space Representations for NLP, 13–18

Liu, P. and H. Meng. 2014. "Seemgo: Conditional random fields labeling and maximum entropy classification for aspect based sentiment analysis". In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland: ACL and Dublin City University, August 2014, pages 527–531.

Mikolov, T., Chen, K.; Corrado, M. and Dean, J. 2013. Efficient Estimation of Word Representations in Vector Space. Proceedings of Workshop at ICLR, 2013.

Nandan, N.; D. Dahlmeier, A. Vij, and N. Malhotra. 2014. "Sapri: A constrained and supervised approach for aspect-based sentiment analysis," in Proceedings of the 8th International Workshop on Semantic Eva-luation (SemEval 2014). Dublin, Ireland: ACL and Dublin City University, August 2014, pages 517–521.

Padró, Ll. and E. Stanilovsky. 2012. FreeLing 3.0: Towards Wider Multi-linguality. In Proceedings of the Language Resources and Evaluation Conference (LREC 2012) ELRA.

Pekar, V.; N. Afzal, and B. Bohnet. 2014. "Ubham: Lexical resources and dependency parsing for aspect-based sentiment analysis," in Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland: ACL and Dublin City University, August 2014, pages 683–687.

Pontiki, M.; D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar. (2014). SemEval 2014 Task 4: Aspect Based Sentiment Analysis. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014) at COLING 2014, pages 27-35. Dublin, Ireland.

Pontiki, M.; D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. AL-Smadi, M. Al-Ayyoub, Y. Zhao, Q. Bing, O. De Clercq, V. Hoste, M. Apidianaki, X. Tannier, N. Loukachevitch, E. Kotelnikov, N. Bel, S. Jiménez-Zafra, G. Eryigit (2016) "SemEval-2016 Task 5: Aspect Based Sentiment Analysis,"Proceedings of the 10th International Workshop on Semantic Evaluation, SemEval '16, San Diego, California, June 16-17, 2016.",19–30,2016, Association for Computational Linguistics.

Schulze Wettendorf, C., R. Jegan, A. Körner, J. Zerche, N. Plotnikova, J. Moreth, T. Schertl, V. Obermeyer, V. Streil, T. Willacker, and S. Evert. 2014. "Snap: A multi-stage xml-pipeline for aspect based sentiment analysis". In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland: ACL and Dublin City University, August 2014, pages 578–584.

Vázquez, Silvia; Muñoz-García, Óscar; Campanella, Inés; Poch, Marc; Fisas, Beatriz; Bel, Nuria; Andreu, Gloria (2014). "A classification of user-generated content into consumer decision journey stages, Neural Networks,58, 68-81.

Veselovská K. and A. Tamchyna. 2014. "U´ fal: Using hand-crafted rules in aspect based sentiment analysis on parsed data". In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland: ACL and Dublin City University, August 2014, pages 694–698.

Vinh, N. X., Epps, J., and Bailey, J. (2009). Information theoretic measures for clustering comparison: is a correction for chance necessary? In Proc. the 26th International Conference on Machine Learning (ICML'09), p.1073- 1080. ACM.

Zhang, F.; Z. Zhang, and M. Lan. 2014. "Ecnu: A combination method and multiple features for aspect extraction and sentiment polarity classification," in Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland: ACL and Dublin City University, pages 252–258.