NoReC: The Norwegian Review Corpus

Erik Velldal, Lilja Øvrelid, Eivind Alexander Bergem, Cathrine Stadsnes, Samia Touileb, Fredrik Jørgensen[†]
Language Technology Group, Department of Informatics, University of Oslo
{erikve, liljao, eivinabe, cathsta, samiat}@ifi.uio.no
[†]Schibsted Media Group
fredrik.jorgensen@schibsted.com

Abstract

This paper presents the Norwegian Review Corpus (NoReC), created for training and evaluating models for document-level sentiment analysis. The full-text reviews have been collected from major Norwegian news sources and cover a range of different domains, including literature, movies, video games, restaurants, music and theater, in addition to product reviews across a range of categories. Each review is labeled with a manually assigned score of 1–6, as provided by the rating of the original author. This first release of the corpus comprises more than 35,000 reviews. It is distributed using the CoNLL-U format, pre-processed using UDPipe, along with a rich set of metadata. The work reported in this paper forms part of the SANT initiative (Sentiment Analysis for Norwegian Text), a project seeking to provide open resources and tools for sentiment analysis and opinion mining for Norwegian.

Keywords: Sentiment Analysis, Opinion Mining, Corpus, Norwegian, Reviews

1. Introduction

For Norwegian, training and evaluation data are still lacking for many core NLP tasks. The current work aims to fill the gap for the particular task of sentiment analysis. The SANT project – Sentiment Analysis for Norwegian Text – seeks to create, and make publicly available, resources and tools for sentiment analysis for Norwegian. The SANT effort described in the current paper marks the release of the Norwegian Review Corpus[1] (NoReC). The dataset comprises more than 35,000 full-text reviews from a range of different domains, collected from several of the major Norwegian news sources. Each review is rated with a numerical score on a scale of 1–6, and can be used for training and evaluating models for document-level sentiment analysis, i.e., the task of predicting overall positive or negative polarity for a given text.

1.1. Rating by dice

A particularity of review journalism in Norway, is the wholesale adoption of dice rolls ('terningkast') as a standard rating scale: The item under review is rated on a scale of 1–6, commonly visualized by the face of a die with a corresponding number of 'dots' or pips. The practice is thought to have been introduced already in 1952 when reviewing movies in the newspaper *Verdens Gang (VG)*. Given that the result of a die roll is otherwise associated with randomness, it is somewhat surprising that it would catch on as a symbol for summarizing reviews – something one would typically hope to construe as a well-informed and deliberate judgment of merit and quite the opposite of chance or luck. Nontheless, by now it has found widespread use in all sorts of arts and consumer journalism and is used when reviewing everything from books, theater and music, to home electronics, restaurants, and children's clothing.

The rating practice described above has several benefits for the goal of document-level SA: (i) It eliminates the need for costly manual annotation since the numerical rating (i.e., the die roll) directly provides us with labels that can be used for training models for detecting the overall document polarity. (ii) There is no need for manually defined mappings to align different rating schemes as the reviews all use a uniform scale. (iii) The wide range of available news sources using the same rating practice, including all the major national newspapers, facilitates the creation of a large-scale dataset. (iv) Models trained on the dataset can be expected to generalize well across domains given the balance of different topics covered in the corpus.

1.2. Sources and partners

The SANT project represents a newly initiated collaboration between the Language Technology Group (LTG) at the Department of Informatics at the University of Oslo, and three of Norway's largest media groups; the Norwegian Broadcasting Corporation (NRK – the state-owned public broadcaster) and the privately held Schibsted Media Group and Aller Media. This first release (Ver. 1.0.1) of NoReC comprises 35,189 reviews extracted from eight different news sources contributed by the three media partners. In terms of publishing date the reviews mainly cover the time span 2003–2017, although it also includes a handful of reviews dating back as far as 1998. We briefly present the sources provided by the different partners below.

Schibsted Media Group The Schibsted group has contributed content from their full portfolio of Norwegian news sources: VG, Aftenposten, Fædrelandsvennen, Bergens Tidende, and Stavanger Aftenblad. While the latter three rank among Norway's largest regional newspapers, Aftenposten is the largest national newspaper in terms of circulation and VG is the largest online news source with more than 2.4 million readers across all platforms.

Aller Media The Aller publishing company has contributed content from two sources. The first is the online version of the newspaper Dagbladet – the second most vis-

[1] `https://github.com/ltgoslo/norec`

Source	Abbr.	# Reviews
Verdens Gang	VG	11,888
Dagbladet	DB	5300
Stavanger Aftenblad	SA	5146
P3.no	P3	5017
DinSide.no	DinSide	2944
Fædrelandsvennen	FVN	2296
Bergens Tidene	BT	1675
Aftenposten	AP	923

Table 1: Number of reviews across sources (also showing abbreviated names).

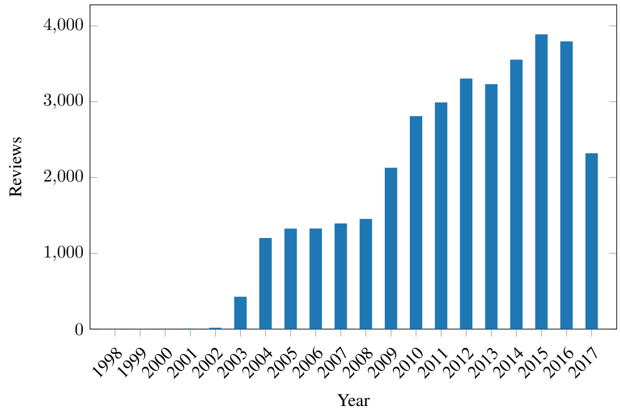

Figure 1: Number of reviews over time.

ited online news source in Norway – publishing reviews for music recordings and live performances, theater and related stage performances, movies, literature, restaurants and more. The second source, DinSide.no, is a website specializing in product reviews, covering a wide range of product types, from home electronics to cars and clothing.

NRK The Norwegian Broadcasting Corporation is a state-owned media house, with a special mandate to be a non-commercial, politically independent public broadcaster. For the review corpus, NRK has contributed content from the website P3.no which has an extensive back catalog of 'die-rated' reviews of movies, TV series, computer games, and music (both recordings and live performances).

2. Corpus creation

The original document collections were provided from the media sources in various JSON, HTML and XML formats, and a substantial effort has gone into identifying relevant documents and extracting text and associated metadata. The extraction process can be summarized by the following four steps: (i) Identify reviews, (ii) convert review content to an intermediate and canonical HTML format, (iii) extract text and pass it through linguistic pre-processing, producing representations in CoNLL-U format, and finally (iv) extract relevant metadata to a JSON representation with normalized attribute–value names. We briefly comment on each of these steps in turn below.

2.1. Identifying reviews

Some of the initial data dumps also included other articles beyond reviews, and in these cases reviews had to be identified. While in some cases this can be done simply by checking for an appropriate metadata field indicating the rating score, other cases require checking for links pointing to an image of a die (indicating the rating), or similar heuristics. Moreover, for some of the sources, a single document may contain multiple reviews, for example for product comparisons. In these cases we had to identify and separate out the different sub-reviews. Different publishing conventions require targeting different types of cues in the document structure, like headers, bold-faced content or die-face images. This also involves extraction of titles and rating scores for the different sub-reviews. The identified sub-reviews become separate documents in the NoReC data set. In total, 35,189 distinct reviews were extracted from the data provided by the media partners. Table 1 shows the number of reviews included from the various sources, while Figure 1 shows the number of reviews per year. Note that because the dataset was assembled in 2017, we only have partial data for that year. As is also evident, the corpus contains only very few reviews prior to 2003.

2.2. Converting content to canonical HTML

The raw data dumps from the sources are mostly in HTML format, but may also be e.g. JSON objects, and have different conventions for document structuring and use of markup. In order to streamline the downstream text extraction, all documents are converted to a 'canonical' HTML format where all textual content is located either inside a header or a paragraph tag. In addition to containing the review text, the raw documents also contain images, ads and other content not considered part of the running text. In order to identify and mark the non-relevant text we use a combination of heuristics based on simple string matching and properties like paragraph length and ratio of content to markup. For example, care was taken to identify 'you-might-also-be-interested-in' type links that are injected throughout the texts in an attempt to keep the reader on the website and generate more clicks. Importantly, however, we chose not to remove any content when converting to our intermediate HTML format, instead introducing a new tag – `remove` – in which we enclose content considered non-relevant. This non-destructive approach preserves the original content, as to not close the door on changes to the subsequent task of text extraction later.

2.3. Linguistic enrichments and CoNLL-U

Given the canonical HTML format described above, it is straightforward to extract the relevant text. In order to enable various types of downstream uses of the dataset, we further pre-process the raw text using the UDPipe toolkit (Straka et al., 2016), representing each review as a CoNLL-U file, following the format defined in Universal Dependencies version 2.[2] In this step we perform sentence segmentation, tokenization, lemmatization, morphological analysis, part-of-speech tagging and dependency parsing, following

[2] http://universaldependencies.org/format.html

the Universal Dependencies scheme (Nivre et al., 2016). However, the pre-processing set-up is slightly complicated by the fact that the Norwegian language has two official written standards – Bokmål (the main variety) and Nynorsk – both of which are represented in the review corpus. Below we first describe how language identification is performed, and then go on to give more details about UDPipe and the resulting CoNLL-U data.

Identifying language varieties The two official varieties of Norwegian are closely related and they are mostly distinguished by minor lexical differences. Still, the differences are strong enough that different pre-processing pipelines must be used for the different standards (Velldal et al., 2017), hence it is important to identify the standard used within a particular document. We use the `langid.py` (Lui and Baldwin, 2012) language identification tool to identify the standard for each review, using its pre-trained models.[3] We performed an evaluation of `langid.py` on 1599 reviews of which 1487 were written in Bokmål and 112 in Nynorsk (based on selecting reviews from authors known to write in a given variety). On this sample `langid.py` achieved 100% accuracy. While the main variety, i.e. Bokmål, dominates the distribution in the corpus with 34,656 documents, we also identified 533 documents in Nynorsk (mainly from the sources Fædrelandsvennen, Bergens Tidende and P3.no).

UDPipe configuration We apply UDPipe (Straka et al., 2016) v.1.2 with its pre-trained models for Norwegian Bokmål and Nynorsk. This version of the UDPipe software and the pre-trained models were developed for the CoNLL 2017 shared task (Zeman et al., 2017), which was devoted to parsing from raw text to Universal Dependencies for more than 40 different languages. We use the models trained for participation in the shared task (Straka and Straková, 2017), not the models provided as baseline models for the participants. The Norwegian models were trained on the UD 2.0 versions of the Norwegian UD treebanks (Øvrelid and Hohle, 2016; Velldal et al., 2017) in conjunction with the aforementioned shared task, and the subsequent choice of model (Bokmål vs Nynorsk) was determined by the language identified for each particular review.

UDPipe obtained competitive results for Norwegian in the shared task, with rankings ranging between first place (lemmatization; both variants) and ninth place (Bokmål dependency parsing LAS) out of 33 participating teams. For reference, in terms of performance for the different subtasks, UDPipe reported F1 scores – for Bokmål / Nynorsk respectively – on sentence segmentation of 96.38 / 92.08, tokenization of 99.79 / 99.93, lemmatization of 96.66 / 96.48, morphological analysis of 95.56 / 95.25, part-of-speech tagging of 96.83 / 96.54, and Labeled Accuracy Scores for dependency parsing of 83.89 / 82.74.

[3] `langid.py` can actually identify three different variants: `no`, `nn` and `nb`, for Norwegian (mixed), Nynorsk and Bokmål, respectively. While the precise details of how the classifier was trained are not clear, it appears to us after some experimentation that the classfication of Bokmål is more accurate when specyfing `no` rather than `nb` and hence is what we use here (together with `nn`). We still use the language codes *nb* and *nn* when adding information about the detected standards to the metadata in NoReC.

	#
Documents	35,189
Sentences	918,681
Tokens	14,998,667
Types – full-form	511,150
Types – lemmas	438,306
Average document length	426

Table 2: Basic corpus counts.

CoNLL-U files When extracting the text from the canonical HTML to pass it to UDPipe, we strip away all mark-up and discard all content marked for removal as described in Section 2.2. Double newlines were inserted between paragraphs and excess whitespace trimmed away. Importantly, however, the text structure is retained in CoNLL-U by taking advantage of the support for comments to mark paragraphs and sentences. In addition to the global document ID number, each paragraph and sentence is also assigned a running ID within the document, using the following form:

- Paragraphs: `<review-id>-<paragraph-id>`, e.g. `000001-03` for paragraph 3 in document 1.
- Sentences: `<review-id>-<paragraph-id>-<sentence-id>`, e.g. `000001-03-02` for sentence 2 in paragraph 3 in document 1.

After completing the UDPipe pre-processing, the corpus comprises a total of 918,681 sentences and 14,998,667 tokens; see Table 2 for an overview of some core corpus counts. A script for executing the entire pipeline from text extraction through UDPipe parsing is made available from the NoReC git repository.

2.4. Metadata and thematic categories

For all the identified reviews, we also provide various kinds of relevant metadata, made available in a JSON representation with normalized attribute–value names across reviews. Metadata extracted from the various sources include information like the URL of the originally published document, numerical rating, publishing date, author list, domain or thematic category, original ID in the source, and more. Beyond this we also add information about the identified language variety (Bokmål/Nynorsk), assigned data split (test/dev/train, as further described in Section 2.5.), assigned document ID, and finally a normalized thematic category.

Thematic categories The 'category' attribute warrants some elaboration. The use of thematic categories and/or tags varies a lot between the different sources, ranging from highly granular categories to umbrella categories encompassing many different domains. Based on the original inventory of categories, each review in NoReC is mapped to one out of nine normalized thematic categories, using English names. The distribution over categories is shown in Table 3, sorted by frequency.

For some sources, this normalization is a matter of simple one-to-one mapping, while for others it is more complex,

Category	#
screen	13,085
music	12,410
literature	3526
products	3120
games	1765
restaurants	534
stage	530
sports	117
misc	102
total	35,189

Table 3: Number of reviews across categories.

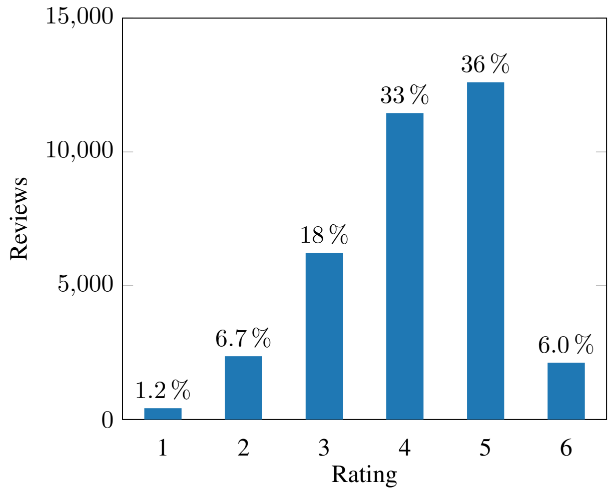

Figure 2: Number of reviews across ratings.

involving heuristics based on the presence of certain tags and keywords in the title. The granularity in the final set of categories is limited by the granularity in the sources. However, the original (Norwegian) source categories are preserved in a seperate attribute in the metadata ('source-category').

As seen from Table 3, the two categories that are by far the largest are 'screen' and 'music'. While the former covers reviews about movies and TV-series, the latter covers both musical recordings and performances. The related category 'stage' covers theater, opera, ballet, musical and other stage performances besides music. The perhaps most diverse category is 'products', which comprises product reviews across a number of sub-categories, ranging from cars and boats to mobile phones and home electronics, in addition to travel and more. The remaining categories of 'literature', 'games', 'restaurants', and 'sports' are self-explanatory, while the 'misc' category was included to cover topics that were infrequent or that could not easily be mapped to any of the other categories by simple heuristics.

2.5. Formats and availability

Distributed under a CC BY-NC 4.0 license,[4] NoReC is available for download from the following git repository:

`https://github.com/ltgoslo/norec`

Formats NoReC is distributed in two formats. The first is the CoNLL-U format as described in Section 2.3., containing tokenized and lemmatized text annotated with PoS tags and dependency graphs. This is considered the primary format. Secondly, we also distribute the canonical HTML representation of the 'raw' review documents as described in Section 2.2. For each format, each review is stored as a separate file, with the filename given by the review ID. To facilitate a low barrier of use for different types of end-users, we also include scripts for converting from CoNLL-U to running tokenized text (using either full-forms or lemmas) and from HTML to raw text without pre-processing. The metadata for each review is provided as a JSON object, all listed in a single file and indexed on the document IDs. The NoReC git repository also includes a Python module with basic functionality for reading the CoNLL-U and JSON representations, as to make experimentation with the corpus as accessible and convenient as possible.

[4] `https://creativecommons.org/licenses/by-nc/4.0/`

Train/dev/test splits To facilitate replicability of experiments the corpus comes with pre-defined standard splits for training, development and testing, with a 80–10–10 ratio. We created the splits per category – sorting the reviews for each category by date and then reserving the first 80% for training, the next 10% for development and the final 10% for testing. This strategy ensures an identical category distribution across splits, while at the same time preserving a stable distribution of ratings as well. The distribution of news source across splits will vary more, but this is less critical as we are primarily interested in ensuring a balanced distribution for ratings and categories. Defining the splits relative to the timeline also reduces the risk of having different reviews of the same item in different splits (e.g., the same product reviewed by multiple news sources). It also generally presents a more realistic test scenario: a trained model will typically be applied to fresh data, with all the time effects that this entails, like shifts in vocabulary, etc.

3. Distribution of ratings

From the perspective of SA, the most immediately relevant piece of metadata in the corpus is obviously the rating. As discussed previously, all the reviews were originally published with an integer-valued rating between 1 and 6, visually indicated using the face of a die. Figure 2 shows the distribution of reviews relative to rating scores. We see that the distribution is highly skewed, with rating values of 4 and 5 being the most common, while very few reviews are given the lowest possible rating of 1. We observe a sharp drop in frequency when moving from rating 5 to 6, perhaps indicating that the distance between these two ratings is perceived as greater than between say 4 and 5.

A similar tendency to lean towards the higher ratings is typically also reported for user-generated reviews, though with a stronger preference for the highest score (Baccianella et al., 2009). In Figures 3 and 4 we see a more detailed view of the rating distribution for each category and source.

In Figure 3 we see that the 'stage' and 'products' categories are most strongly skewed towards the rating of 5. As most of the product reviews were gathered from 'DinSide.no', we see a similar distribution for this source in Figure 4.

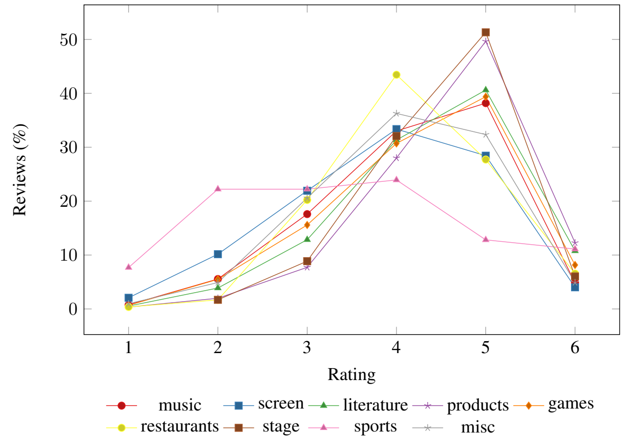

Figure 3: Distribution of ratings for each category.

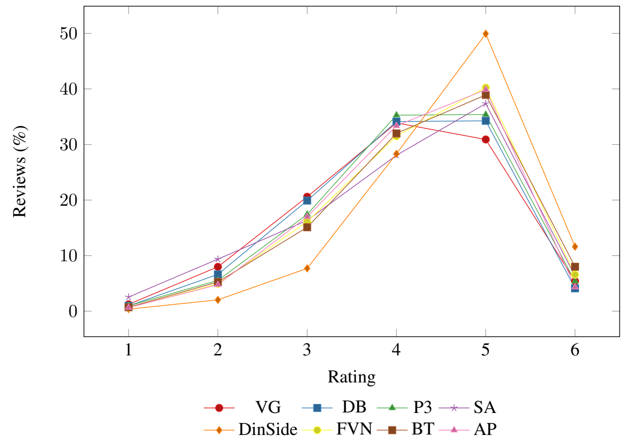

Figure 4: Distribution of ratings for each source.

Overall, however, most source and categories exhibit a similar pattern; a majority of ratings at 4 and 5, and with relatively fewer reviews in either extreme of the scale. One notable exception is the category 'sports', and to some degree also 'restaurants'.

4. Related work

The dataset described in the current paper is the first of its kind for Norwegian. For other languages, however, the field has seen a substantial amount of SA research based on rated reviews, either user-generated or by professional reviewers. This has often been based on single-domain datasets, and examples include (for English unless otherwise noted) movie reviews collected from aggregator sites like IMDb.com (Pang and Lee, 2004; Maas et al., 2011) and RottenTomatoes.com (Pang and Lee, 2005; Socher et al., 2013), hotel reviews from TripAdvisor (Wang et al., 2010), book reviews (in Arabic) (Aly and Atiya, 2013), app reviews compiled from Apple App store and Google Play (Guzman and Maalej, 2014), and reviews of restaurants and other businesses in the Yelp open dataset.[5] However, the unbalanced nature of these datasets (single domains) can impose inherent limitations on the ability of models to generalize. Some datasets combine reviews from multiple domains for better balance, like the French SA corpus of Vincent and Winterstein (2013), combining reviews of movies, books and hotels (from Allocine.fr, Amazon.fr, and TripAdvisor.fr, respectively), or the Arabic SA corpus of ElSahar and El-Beltagy (2015), combining reviews of hotels, restaurants, movies, restaurants and product reviews (from TripAdvisor, elCinema.com, Qaym.com and Souq.com). There also exists several datasets based on product reviews from Amazon, which can potentially also have the advantage of covering a more diverse selection of domains. An example includes the Amazon dataset of Blitzer et al. (2007), comprising reviews of books, DVDs, electronics, and kitchen appliances.

The NoReC dataset described in the current paper covers a wide range of domains, combining product reviews across a diverse range of categories, such as literature, restaurants, sports, music, various stage arts and more.

5. Future work

While we plan to further add more reviews to NoReC for future releases, covering additional domains or categories, the SANT project will also seek to build on NoReC to (i) experiment with both polarity classification and rating inference on the document-level using neural architectures, (ii) extract SA lexicons encoding the polarity of individual words, and finally (iii) also move beyond the document-level and manually add more fine-grained and aspect-based SA annotations for a sub-set of the corpus. These annotations will also be used to (iv) train a classifier separating subjective and objective sentences. Across all these activities, the various thematic categories will be useful for assessing cross-domain effects (e.g., how well does an SA classifier trained on movie reviews perform for home electronics?) and potentially even for training domain-specific models. It will of course also be important to assess how well SA resources developed on the basis of the reviews generalize to non-review texts, and we plan to annotate aspect-based sentiment for a selection of general-domain news texts as well.

6. Summary and outlook

The current paper has described the creation of the Norwegian Review Corpus; NoReC (Ver. 1.0.1). The final dataset comprises more than 35,000 full-text reviews ($\approx$ 15 million tokens) from a wide range of different domains, collected from several major Norwegian news sources. Each review is rated with a numerical score on a scale of 1–6, and can be used for training and evaluating models for document-level sentiment analysis. Resources for sentiment analysis have so far been unavailable for Norwegian. While the primary distribution format of the corpus is CoNLL-U – based on only the extracted text and applying UDPipe for a full preprocessing pipeline from sentence segmentation to dependency parsing – the release also includes HTML representations of the full reviews with all content preserved. Each review is in addition associated with a rich set of metadata, including thematic category. We also provide pre-defined splits for training, development and testing.

[5] `https://www.yelp.com/dataset`

7. Acknowledgments

This work was carried out as part of the SANT research project (Sentiment Analysis for Norwegian Text), funded by an IKTPLUSS grant from the Research Council of Norway (project no. 270908). SANT represents a newly initiated collaboration between the Department of Informatics at the University of Oslo, and three of Norway's largest media groups; the Norwegian Broadcasting Corporation, Schibsted Media Group and Aller Media.

8. Bibliographical references

Aly, M. and Atiya, A. (2013). LABR: A large scale arabic book reviews dataset. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, volume 2 (short papers), pages 494–498, Sofia, Bulgaria.

Baccianella, S., Esuli, A., and Sebastiani, F. (2009). Evaluation measures for ordinal regression. In *Proceedings of the 9th International Conference on Intelligent Systems Design and Applications*, pages 283–287, Pisa, Italy.

Blitzer, J., Dredze, M., and Pereira, F. (2007). Biographies, Bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *Proceedings of the 45th Meeting of the Association for Computational Linguistics*, pages 187–205, Prague, Czech Republic.

ElSahar, H. and El-Beltagy, S. R. (2015). Building large arabic multi-domain resources for sentiment analysis. In *Proceedings of the 16th International Conference on Intelligent Text Processing and Computational Linguistics*, pages 23–34, Cairo, Egypt.

Guzman, E. and Maalej, W. (2014). How do users like this feature? A fine grained sentiment analysis of app reviews. In *Proceedings of the 22nd IEEE International Requirements Engineering Conference*, pages 153–162, Karlskrona, Sweden.

Lui, M. and Baldwin, T. (2012). langid.py: An off-the-shelf language identification tool. In *Proceedings of the 50th Meeting of the Association for Computational Linguistics System Demonstrations*, pages 25–30, Jeju, Republic of Korea.

Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., and Potts, C. (2011). Learning word vectors for sentiment analysis. In *Proceedings of the 49th Meeting of the Association for Computational Linguistics*, pages 142–150, Portland, OR, USA.

Nivre, J., de Marneffe, M.-C., Ginter, F., Goldberg, Y., Hajič, J., Manning, C. D., McDonald, R., Petrov, S., Pyysalo, S., Silveira, N., Tsarfaty, R., and Zeman, D. (2016). Universal dependencies v1: A multilingual treebank collection. In *Proceedings of the 10th International Conference on Language Resources and Evaluation*, pages 1659–1666, Portorož, Slovenia.

Øvrelid, L. and Hohle, P. (2016). Norwegian Universal Dependencies. In *Proceedings of the 10th International Conference on Language Resources and Evaluation*, pages 1579–1585, Portorož, Slovenia.

Pang, B. and Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics*, pages 271–278, Barcelona, Spain.

Pang, B. and Lee, L. (2005). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 43rd Meeting of the Association for Computational Linguistics*, pages 115–124, Ann Arbor, MI, USA.

Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C., Ng, A., and Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1631–1642, Seattle, WA, USA.

Straka, M. and Straková, J. (2017). Tokenizing, POS tagging, lemmatizing and parsing UD 2.0 with UDPipe. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 88–99, Vancouver, Canada.

Straka, M., Hajič, J., and Straková, J. (2016). UDPipe: trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, pos tagging and parsing. In *Proceedings of the 10th International Conference on Language Resources and Evaluation*, pages 4290–4297, Portorož, Slovenia.

Velldal, E., Øvrelid, L., and Hohle, P. (2017). Joint UD parsing of Norwegian Bokmål and Nynorsk. In *Proceedings of the 11th Nordic Conference of Computational Linguistics*, pages 1–10, Gothenburg, Sweden.

Vincent, M. and Winterstein, G. (2013). Construction et exploitation d'un corpus français pour l'analyse de sentiment (building and exploiting a French corpus for sentiment analysis). In *Proceedings of the 20th Conference Traitement Automatique des Langues Naturelles*, volume 2 (short papers), pages 764–771, Les Sables-d'Olonne.

Wang, H., Lu, Y., and Zhai, C. (2010). Latent aspect rating analysis on review text data: A rating regression approach. In *Proceedings of the 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 783–792, Washington DC, DC, USA.

Zeman, D., Popel, M., Straka, M., Hajic, J., Nivre, J., Ginter, F., Luotolahti, J., Pyysalo, S., Petrov, S., Potthast, M., Tyers, F., Badmaeva, E., Gokirmak, M., Nedoluzhko, A., Cinkova, S., Hajic jr., J., Hlavacova, J., Kettnerová, V., Uresova, Z., Kanerva, J., Ojala, S., Missilä, A., Manning, C. D., Schuster, S., Reddy, S., Taji, D., Habash, N., Leung, H., de Marneffe, M.-C., Sanguinetti, M., Simi, M., Kanayama, H., dePaiva, V., Droganova, K., Martínez Alonso, H., Çöltekin, c., Sulubacak, U., Uszkoreit, H., Macketanz, V., Burchardt, A., Harris, K., Marheinecke, K., Rehm, G., Kayadelen, T., Attia, M., Elkahky, A., Yu, Z., Pitler, E., Lertpradit, S., Mandl, M., Kirchner, J., Alcalde, H. F., Strnadová, J., Banerjee, E., Manurung, R., Stella, A., Shimada, A., Kwak, S., Mendonca, G., Lando, T., Nitisaroj, R., and Li, J. (2017). Conll 2017 shared task: Multilingual parsing from raw text to universal dependencies. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 1–19, Vancouver, Canada.