Tracking Changes in ESG Representation: Initial Investigations in UK Annual Reports

**Matthew Purver,[*‡] Matej Martinc,[*] Riste Ichev,[†] Igor Lončarski,[†]
Katarina Sitar Šuštar,[†] Aljoša Valentinčič,[†] Senja Pollak[*]**

[*]Department of Knowledge Technologies, Jožef Stefan Institute, Ljubljana, Slovenia
[†]School of Economics and Business, University of Ljubljana, Ljubljana, Slovenia
[‡]School of Electronic Engineering and Computer Science, Queen Mary University of London, UK
{matthew.purver, matej.martinc}@ijs.si, {riste.ichev, igor.loncarski, katarina.sitar, aljosa.valentincic}@ef.uni-lj.si, senja.pollak@ijs.si

Abstract

We describe initial work into analysing the language used around environmental, social and governance (ESG) issues in UK company annual reports. We collect a dataset of annual reports from UK FTSE350 companies over the years 2012-2019; separately, we define a categorized list of core ESG terms (single words and multi-word expressions) by combining existing lists with manual annotation. We then show that this list can be used to analyse the changes in ESG language in the dataset over time, via a combination of language modelling and distributional modelling via contextual word embeddings. Initial findings show that while ESG discussion in annual reports is becoming significantly more likely over time, the increase varies with category and with individual terms, and that some terms show noticeable changes in usage.

Keywords: environmental, social, governance, ESG, diachronic analysis

1. Introduction

Companies and investors are becoming increasingly aware of the importance of Corporate Social Responsibility (CSR) in their actions, tracking and reporting their impact on society and the environment. One way to examine a company's behaviour in this area is via Environmental, Social and Governance (ESG) criteria. ESG criteria cover a company's environmental impact (Environmental), their relationships with their community including employees, suppliers and customers (Social), and their leadership structures including executive pay, shareholder rights, audits and controls (Governance). ESG analyses are currently performed manually by experts; for example, Lydenberg et al. (2010) define a method for identifying key sustainability performance indicators which requires six detailed analysis steps. Our interest is in developing NLP technologies to help automate this process, to characterise companies in terms of different ESG criteria and understand how these relate to company performance, risk and outlook over time, as well as more general changes in the economic and regulatory environment. Given the increasing investor interest, coupled with the regulatory push in terms of non-financial reporting, mostly driven by sustainability motives, understanding of the connection between corporate ESG reporting and the measurement of ESG is becoming very important. So far, regulatory requirements regarding ESG reporting have been relatively loose. However, this is starting to change very quickly and dramatically, as exemplified by the introduction of the EU taxonomy for sustainable activities, sustainability reporting standards such as SASB, and the most recent evolution in IFRS reporting. While NLP techniques have been developed for specific aspects relating to ESG, particularly environmental concerns (Armbrust et al., 2020) and more specifically, discussion relating to climate change (Luccioni et al., 2020), a more general model for tracking and characterizing ESG reporting has yet to be produced.

In this paper, we outline our initial steps in this direction, by defining a categorised set of 93 ESG terms covering 5 core ESG areas, based on a number of existing resources and filtered by multiple annotators, that can be used to analyse changes in reporting. By assembling a collection of company annual reports and applying analyses based on language modelling and on distributional methods, we show that these terms have the potential to reveal changes in the frequency and in the usage of the language of ESG.

2. Dataset and ESG analysis terms

2.1. Data and pre-processing

We base our analysis on annual reports from FTSE350 companies over the years 2012-2019. To establish a fixed list of companies for comparison purposes, we used the FTSE350 list as of 25th April 2020.[1] Reports were obtained from the publicly accessible collection at www.annualreports.com. Not all companies' reports were available, and to disambiguate between companies with the same ticker on different exchanges, we used only those with reports shown at the London Stock Exchange (LSE); the number of reports obtained, together with total word counts (before preprocessing or

[1]`https://en.wikipedia.org/w/index.php?title=FTSE_350_Index&oldid=953125037`

Year	# Reports	# Words
2012	178	12.5M
2013	181	14.0M
2014	184	15.0M
2015	196	16.3M
2016	198	17.5M
2017	200	18.4M
2018	200	19.6M
2019	202	21.2M
total	1539	134.6M

Table 1: Number of annual reports retrieved by year

tokenization), are shown in Table 1. The reports are published as PDF documents; these were converted to raw text using the `pdf2txt` tool.[2] We tokenize into words and build ngrams of length 1-4 padded with sentence start and end symbols, using NLTK's standard preprocessing tools.[3] While we do not have the rights to redistribute this data, it comes from public sources, and the details required to re-create the dataset are available publicly at `osf.io/rqgp4`.

2.2. ESG term extraction

As our interest is in comparing the ways in which ESG concepts are written about, our first task was to define a set of suitable terms (words or multi-word expressions) for subsequent use in analysing the report text.

Initial seed terms We started with three existing lists of terms likely to relate to ESG concepts, derived from (a) the SASB standards, (b) Schroders and (c) our own work. The first list comes from the 2017 Conceptual Framework for sustainability in accounting set out by the SASB (Sustainability Accounting Standards Board, now part of the Value Reporting Foundation - see www.sasb.org): specifically, we take the "SASB Universe of Sustainability Issues" which defines 5 sub-areas (*environment, social capital, human capital, business model and innovation, leadership and governance*) and gives 4-7 major concepts for each one (SASB, 2017). For example, the concepts for *environment* include *GHG emissions, Air quality, Energy management, Fuel management, Water and wastewater management, Waste and hazardous materials management, Biodiversity impacts*. This list contains 36 terms. Our second source is the Schroders brochure "Understanding sustainable investment and ESG terms" (Schroders, 2021), used to establish "the landscape of activities, strategies that fall under the broad umbrella of ESG and sustainability", and the terms most commonly associated with each. This approach takes the point of view of the investor, and defines 6 sub-areas (*integration, governance & active ownership, screened investments, thematic investing, impact investing, industry organisations and initiatives*), each specified with 7-15 major concepts. For example, the concepts for *thematic investing* include *carbon footprint, climate risk, green investing, renewable energy*. This list contains 62 terms.

Our third and final source is an annotated dataset developed during our own project and described in (Stepišnik-Perdih et al., 2022). For this dataset, sentences were extracted from annual reports of companies listed on US or UK stock exchanges that cover the period 2017 to 2019. Annotation was then performed at the sentence level, with sentences marked as ESG-related or not. Thirteen annotators were used, with each annotator given 500 sentences for annotation. Annotators were second-year graduate students of the MSc in Quantitative Finance and Actuarial Sciences at the School of Economics and Business, University of Ljubljana. Given their field and length of studies, we believe they were well suited to the task of annotating financial texts. Annotators were asked to annotate each of the sentences according to several criteria. First, whether the sentence is relevant from the perspective of corporate business. Second, whether the sentence conveys positive/negative/neutral financial sentiment. Third, whether the sentence expresses an opinion (subjectivity) or states the facts (objectivity). Fourth, whether it is forward-looking or not. Finally, whether it relates to ESG or not. Full details are given in (Stepišnik-Perdih et al., 2022); in this work, we use only the labels with regards to ESG, with a binary label positive (ESG-related) or negative (not ESG-related). The dataset contains 6,500 sentences, within which 24.8% (1,617 sentences) are ESG-related. Based on these annotated sentences, we estimated two 1-2-gram language models using maximum likelihood estimation (using NLTK's standard language modelling tools),[4] one for ESG-related text and one for non-ESG-related text. We then extracted characteristic terms as single words or two-word terms matching the part-of-speech patterns JJ-NN* or NN*-NN* for which the ratio of the language model probabilities $p_{ESG}/p_{nonESG} > 5.0$. This list includes terms concerning a range of ESG aspects, including the environment (*greenhouse gas, meteorological parameters, ambient temperature*), social issues (*female, women, gender pay, human rights, young people, mental health*) and overall standards and reporting concepts (*ethical standard, zero harm, cultural fit, diversity policy*). This list contains 109 single-word and 233 two-word terms.

Term selection We combined these lists to give 440 candidate ESG-related terms of length 1 to 3 words. We randomly shuffled this list, and 4 annotators with finance expertise were independently asked to label each as to whether it was likely to be a representative ESG term, and if so to categorize it according to a 6-way

[2] `https://github.com/pdfminer/pdfminer.six/blob/develop/tools/pdf2txt.py`

[3] `https://www.nltk.org/api/nltk.lm.preprocessing.html`

[4] `https://www.nltk.org/api/nltk.lm.html`

1SC	social capital
2HC	human capital
3BMI	business model & innovation
4LG	leadership & governance
5E	environment
6ESG	environmental social governance

Table 2: ESG category labels, derived from the SASB Conceptual Framework (SASB, 2017)

schema shown in Table 2. The schema consists of the 5 sub-areas of ESG defined by the SASB conceptual framework (SASB, 2017), together with a sixth general category for terms that could not be categorized under any of those 5.
Inter-annotator agreement over the entire list of candidate terms was reasonable, with overall average pairwise Cohen's kappa 0.50 (minimum 0.32, maximum 0.60). Of the 440 candidate terms, 311 were given a label by at least one annotator, but only 93 were given a label by all four annotators (i.e. unanimously agreed to be representative ESG terms). We take these 93 terms as our term list for analysis. Over this set, agreement on the ESG category labels was good, with average pairwise Cohen's kappa 0.71 (minimum 0.63, maximum 0.78). We take the most frequently assigned label as the gold-standard ESG category for each term. The final term list is available publicly at `osf.io/rqgp4`.

3. Language modelling analysis

After text pre-processing, we build a language model for each year in our dataset for word ngrams length 1-4, using maximum likelihood estimation (again using NLTK's standard language modelling tools). This allows us to perform comparison across years of the probabilities of occurrence of 1-to-4-word terms, and of the most likely context following occurrences of those terms. To find terms which have changed most in their probability of use, we find the gradient over time: taking the probability of use of a term over time, we apply standard zero-mean/unit-variance scaling, fit a simple linear regression model and extract the first coefficient. We also do the same for the mean probabilities over the set of terms for each ESG category.

ESG categories We find that overall, ESG terms are becoming more likely in company reports over time, in particular since 2015/2016 annual reports: for all the 6 ESG categories, gradients are positive - see Figure 1. However, there are significant differences in the gradients, with some categories growing faster than others. The fastest-growing is *2HC human capital*, followed by *5E environment* and *3BMI business model*; the slowest-growing are *1SC social capital* and the general/other category *6ESG*.

ESG terms Individual ESG terms, on the other hand, vary widely. Some are increasing noticeably in probability, and the 10 most increasing terms include terms

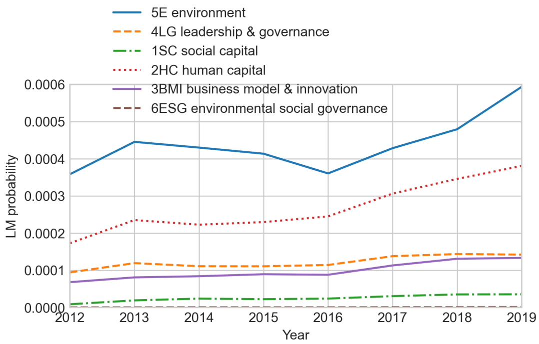


Figure 1: Average probability of mention of ESG term categories over time

from all 5 of our main ESG categories: in 1SC, *human rights*; in 2HC, *talent, wellbeing, pay gap, gender pay*; in 3BMI, *innovation*; in 4LG, *ethical, governance framework*; in 5E, *climate change, renewable*. Figure 2 shows these 10 most increasing terms. Similar to the findings related to ESG categories, terms associated with 2HC human capital exhibit some of the strongest growth in probability after 2015/2016. It also seems that the driver in the growth of 5E environment category is mostly related to the term *renewable*.

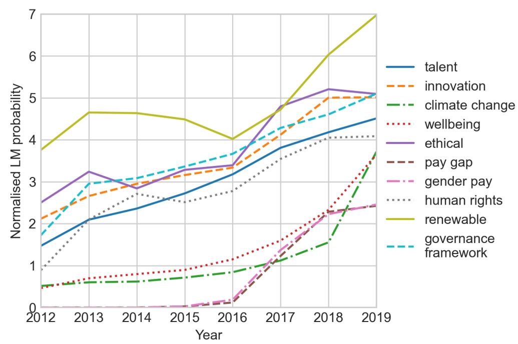


Figure 2: Probability of mention of 10 most increasing ESG term categories over time

Many other terms, though, are increasing more slowly, and some are decreasing in probability. Figure 3 shows the 10 most decreasing terms: *compensation, corporate responsibility, environmental management, waste management, pension plan, water treatment, human resources, emission control, compliance committee, business ethics*. Again, most categories are represented (with the exception of 1SC, social capital), but the nature of the terms is different. In 2HC (human capital), the emphasis now seems less on equality (*pay gap, gender pay*) and on individuals (*talent, wellbeing*) and more on general issues (*human resources*) and on financial aspects (*compensation, pension plan*). In 5E (environment), the emphasis now seems less on specific issues and more on policies and compliance. This

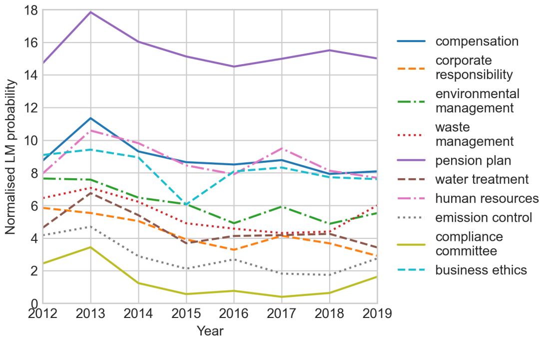


Figure 3: Probability of mention of 10 most decreasing ESG term categories over time

therefore seems to have potential to reveal some finer-grained changes over time in the discussion of ESG and the emphasis placed on certain aspects; however, it seems likely that our 6 general ESG categories — and therefore taxonomies such as the SASB Conceptual Framework from which they were taken — may not be fine-grained enough to analyse this quantitatively, and could benefit from more detailed subcategorization to allow direct analysis.

4. Contextual analysis

Given this, we next turn to look at whether these terms have changed in usage over time, as well as frequency: changing likelihood of use of a term may simply indicate a straightforward change in its frequency of use in reporting, but may also be associated with changes in the context of its use, as it becomes used in different ways or with different emphases. One possible way to examine this is again through language modelling, by inspecting changes in the most likely continuations predicted by a language model after observing a term.[5] However, for the terms of interest here, we find few differences: likely continuations are dominated by syntactic dependencies and end-of-sentence predictions.

Instead, we applied a distributional method used in our previous work to examine diachronic changes in word usage (Montariol et al., 2021). For each word, we generate a set of contextual word embeddings using BERT (Devlin et al., 2019), summing over sub-word tokens where required. These vector representations are then clustered using k-means (taking the clusters to approximate fine-grained word senses), and the resulting cluster distributions compared across years. We measure distance between distributions using Jensen-Shannon divergence (JSD) (Lin, 1991), and take this as a measure of the relative degree of usage shift.

Overall degree of change This allows us to rank our ESG terms by their degree of usage shift over time. The most changing terms (those with the biggest overall distance between distributions from 2012 to 2019) include many terms whose likelihood increased most in the analysis of the previous section (e.g. *wellbeing, talent*); as well as other terms with only moderate likelihood increase (e.g. *pollution, greenhouse gas*). Interestingly, though, some of the terms whose usage changed least (those with the smallest overall distance from 2012 to 2019) also include terms whose likelihood increased sharply (e.g. *innovation, human rights*). Figure 4 shows how JSD varies across time for some selected terms which show high degrees of usage change (*wellbeing, talent, pollution, greenhouse gas*) together with some which show low change (*environmental, innovation*). Erratic and significant year-to-year changes for individual terms might be idiosyncratic. For example, a significant increase for the term *pollution* between 2013 and 2015 might be related to a particular industry and/or single major environmental disaster. The usage of the term *innovation* changed by a factor of 10 less than the usage of the term *wellbeing*; although the likelihood increase (probability gradient) of *innovation* was 50% higher than that of *wellbeing*.

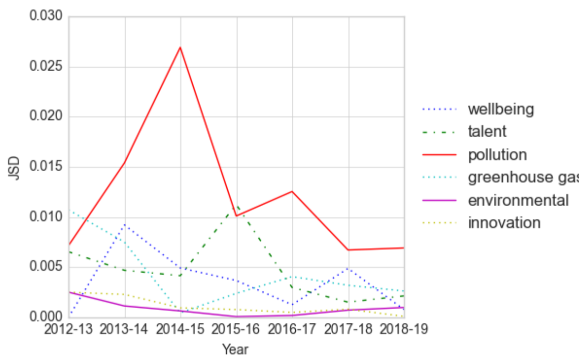
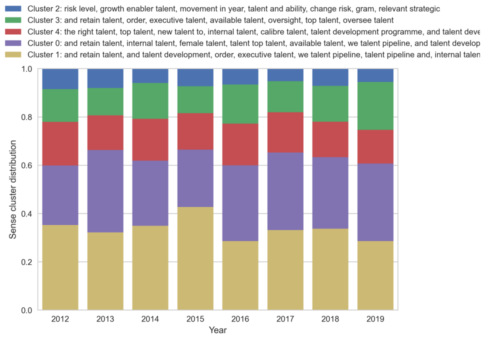


Figure 4: Jensen-Shannon divergences between adjacent year pairs over time, for selected high-change and low-change terms

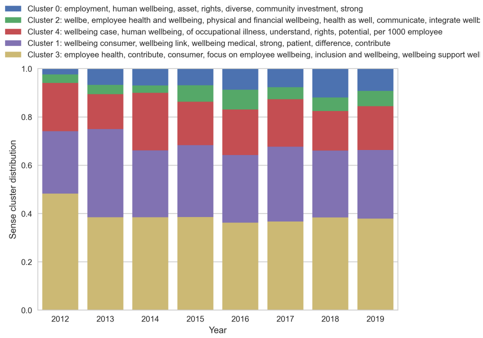
Cluster analysis For terms which show high usage change, this raises the question of in what ways the usage has changed. A full investigation must be approached qualitatively, in order to understand what themes are emerging or being reduced; but we can gain some insight by examining which of the sense clusters become more or less frequent. Figure 5 shows the cluster distributions (proportions of sentences assigned to each cluster by k-means) over time for two high-change terms (*talent, wellbeing*) and two low-change terms (*environmental, innovation*). For each cluster, we show a set of representative keywords: these are extracted by finding those with the highest tf-idf score when considering a cluster as a single document and all clusters as a corpus, while excluding stopwords and words appearing in over 80% of clusters.

Taking one example, we can inspect the term *wellbe-*

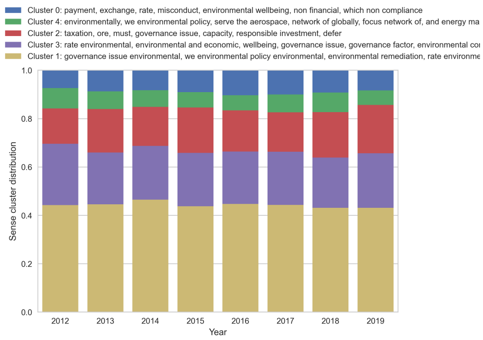
[5] Manual analysis and coding may allow deeper insights in future when time allows, see e.g. (Burgers and Ahrens, 2018).



(a) *talent*



(b) *wellbeing*



(c) *environmental*

(d) *innovation*

Figure 5: Cluster distributions over time for four selected ESG terms: (a),(b) show two terms whose distributions change most, and (c),(d) show two terms whose distributions change least

ing (Figure 5(b)), and see that sense cluster 3 decreases noticeably in likelihood over time, while clusters 0 and 2 increase. The extracted keywords themselves give some limited insight into the differences: the increasing clusters 0 and 2 include keywords relating to *diversity, community* and *financial wellbeing*; however there is also a significant amount of overlap, with *employee wellbeing* and *employee health* seemingly covered in both increasing and decreasing senses. Manual inspection of sentences assigned to particular clusters gives some more insight, with sense cluster 3 (decreasing) seeming to be more focused on general statements of values, while clusters 0 and 2 (increasing) are more specific. Cluster 0 contains a high proportion of concrete statements of past actions, while cluster 2 contains more focused statements about health and financial aspects of wellbeing. Table 3 shows some (manually chosen) examples.

Acknowledgements

The authors acknowledge the financial support from the Slovenian Research Agency for research core funding (No. P2-0103 and No. P5-0161), as well as for funding of the research project *Quantitative and qualitative analysis of the unregulated corporate financial reporting* (No. J5-2554).

5. Bibliographical References

Armbrust, F., Schäfer, H., and Klinger, R. (2020). A computational analysis of financial and environmental narratives within financial reports and its value for investors. In *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*, pages 181–194.

Burgers, C. and Ahrens, K. (2018). Change in Metaphorical Framing: Metaphors of TRade in 225 Years of State of the Union Addresses (1790–2014). *Applied Linguistics*, 41(2):260–279, 12.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186.

Lin, J. (1991). Divergence measures based on the shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151.

Luccioni, A., Baylor, E., and Duchêne, N. A. (2020). Analyzing sustainability reports using natural language processing. *ArXiv*, abs/2011.08073.

Lydenberg, S., Rogers, J., and Wood, D. (2010). From transparency to performance: Industry-based sustainability reporting on key issues. Technical report, Hauser Center for Nonprofit Organizations at Harvard University. Available

0	Health and wellbeing: During the year, the Committee reviewed the significant amount of work being undertaken across the Group as we continue to promote, support and deliver a multitude of health and wellbeing activities for employees, comprising a mix of physical, mental and occupational services.
0	As a Board, we are satisfied that there is no complacency in the business with regards to health and safety but we will continue to challenge the leadership team to maintain a constant focus on the safety of our colleagues and customers and their health and wellbeing, particularly in areas where some risk inevitably arises such as driving within our predominantly route based businesses, and working at height.
0	As a result, over 90% of leaders feel that they are comfortable having conversations about mental health with their team peers and managers, know about and are comfortable signposting colleagues to the resources available to them • 4,000 colleagues have taken advantage of the free access to Headspace offered, collectively completing over 66,000 sessions since launch Physical wellbeing • 8% of colleagues have taken advantage of our discounted fitness proposition which launched in 2018, and they've certainly been active, clocking up over 31,000 gym visits, and the equivalent of 970 days worth of exercise!
0	The year also saw us launch a Wellbeing Programme encouraging open dialogue through monthly presentations on a range of health topics including healthy eating, drugs awareness, emotional well-being and cancer; making sure our people are aware of additional supporting information and the free health and wellness resources available such as flu jabs, eye tests and general physical wellbeing checks.
2	Health and wellbeing initiatives have been selected locally and include well person clinics, office fruit baskets and exercise classes.
2	We have health and wellbeing champions across the business globally and this year they have organised and promoted a range of health and wellbeing activities in our offices, from informative briefing sessions on healthy living through to massage sessions.
2	Nearly four in five employees believe that [ANON] values their health and wellbeing, up nine percentage points in 2017 alone following the launch of a highly successful Health and Wellbeing programme.
2	Financial and health wellbeing is top of employer agendas and we continue to support them and their employees with further development of [ANON] Wellbeing, a set of services aimed at helping employers build healthier, happier and more productive workforces.
3	By inspiring and enabling people to never stop growing and take charge of their wellbeing Unlock capacity for growth
3	Our purpose statement, which was developed in partnership with colleagues from across the business is to be the local partner taking care of journeys that enhance the lives and wellbeing of our communities across the world.
3	Cultivating community spirit and wellbeing in [ANON] Middle-East with [ANON] Sports.
3	The wellbeing of everyone who interacts with our business is a top priority for [ANON].
3	Our beautiful homes, passionate people and excellent wellbeing services set the scene for the creation our communities.

Table 3: Example sentences with k-means cluster labels for the term *wellbeing*

from https://iri.hks.harvard.edu/links/transparency-performance-industry-based-sustainability-reporting-key-issues.

Montariol, S., Martinc, M., and Pivovarova, L. (2021). Scalable and interpretable semantic change detection. In *Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4642–4652.

SASB. (2017). SASB conceptual framework. Technical report, Sustainability Accounting Standards Board, San Francisco, CA, February. Available from https://www.sasb.org/wp-content/uploads/2020/02/SASB_Conceptual-Framework_WATERMARK.pdf.

Schroders. (2021). Understanding sustainable investment and ESG terms. Technical report, Schroder Investment Management Limited, London, UK. Available from https://www.schroders.com/de/sysglobalassets/global-assets/english/campaign/sustainability/interpret/understanding-sustainable-investment-and-esg-terms.pdf.

Stepišnik-Perdih, T., Pelicon, A., Škrlj, B., Žnidaršič, M., Lončarski, I., and Pollak, S. (2022). Sentiment classification by incorporating background knowledge from financial ontologies. In *Proceedings of the 4th Financial Narrative Processing Workshop*.