It does not rain more in the UK – A corpus analysis comparing the optimism variable in Swedish and UK-based listed companies' sustainability reports

Lina de Neergaard (Erenvidh)

Department of English

Independent Degree Project at Undergraduate Level
English Linguistics

Spring 2023

Supervisor: Kathrin Kaufhold

Stockholms universitet

Abstract

This thesis sets out to compare the level of optimism, quantified as the optimism variable, in UK and Swedish-based corporate Environmental, social, and corporate governance ("ESG") disclosures published in English. The comparison is conducted by creating two corpora, one UK-based and one Swedish-based. The corpora are made up by 21 ESG disclosures per country, all collected from financial service companies and published between 2019 and 2022. The two corpora are then semantically analysed, using the text-mining software DICTION. DICTION compares the language and lexemes used in each ESG disclosure against a dictionary with pre-assigned values per word and strings of words, creating an optimism score for each ESG disclosure. These scorings are aggregated per country, creating a mean and a standard deviation profile for the UK-based and the Swedish-based companies, respectively.

By applying a T-test, comparing the means of the two populations regarding the optimism variable it can be concluded with 95 % confidence that there is no difference. It thus seems that the Swedish-based and the UK-based are similar in their level of language optimism.

Keywords

Corpus analysis, semantics and Computer assisted discourse analysis.

Contents

1. Introduction

As society becomes increasingly aware of environmental, social, and governance ("ESG") challenges, investors and stakeholders are seeking greater transparency from companies. Almost 15 years ago Bahita (2008) argued that while linguists can handle the written aspects of certain professional genres, they may not fully understand the communication norms of the professional world. This was exemplified in their paper by the corporate disclosures of Enron. This now defaulted energy company manipulated their bookkeeping to inflate publicly disclosed earnings and drive stock prices. It used and abused linguistic and rhetorical strategies to further its stock market communication (Ibid). Jones, Gaia and Aresu (2020) found that European listed banks reacted to the global financial crisis in their annual reports by applying two reporting scenarios: reactive impression management and retrospective sense-making. Bondi and Yu (2019) found that companies tend to orchestrate voices into symphony in their ESG disclosure, with most sources represented as individuals with specified names. Their analysis shows that companies from different cultural backgrounds have different preferences in selecting and representing sources, with Italian and American reports favouring managers' voices and Chinese reports showing a preference for voices from employees and clients (Ibid). Similarly, Laskin and Nesova (2022) found that American companies tend to be overly positive in their ESG reporting compared to the rest of their financial reporting. Albitar, Abdoush, and Hussainey (2021) discovered significant differences in interpretation between UK and German firms when applying the same accounting concepts. This indicates that there may be differences in how companies disclose information based on the topic, as well as differences in interpretation of the same legal framework among companies in different territories.

According to Laskin and Nesova (2022), Sustainability as a concept is generally traced back to a 1987 United Nations report titled "Our Common Future". ESG reporting has since become formalized as part of annual financial disclosure (Ibid). The European Union has implemented ESG-related regulations such as the Non-Financial Reporting Directive (NFRD) and the Sustainable Finance Disclosure Regulation (SFDR). Publicly traded companies, whose shares are bought and sold on public markets like the London Stock Exchange or the Swedish OMX Nasdaq, are subject to special information disclosure rules, including ESG disclosures. Doshi, Dowell and Toffel (2013) describe mandatory information disclosure as a regulative mechanism meant to make companies reveal how they operate.

Financial disclosure in the UK and Sweden is governed by various regulations and standards, including the Companies Act of 2006 and the Listing Rules of the Financial Conduct Authority (FCA) in the UK, and the Swedish Companies Act, the Swedish Accounting Act, and the EU's NFRD in Sweden. These rules require companies to disclose a wide range of financial and non-financial information, including sustainability and ESG issues. Numerous legislation in EU and non-EU countries mandate publicly traded companies to provide ESG metrics and narratives in both local languages and English, creating a pool of English-written texts issued by companies in different territories.

Differences in the tone of ESG disclosures may be influenced by the legal and regulatory frameworks in Sweden and the UK. The Swedish system is based on common law and has a more prescriptive approach, with laws and regulations mandating ESG reporting and disclosure. In contrast, the UK has a more principles-based approach, relying on voluntary reporting and disclosure based on its tradition of case law. According to Dainow, this is commonly the outcome of case vs. common law jurisdictions (1966). This could result in Swedish companies appearing more optimistic in their disclosures, while UK companies may be more cautious due to the voluntary nature of ESG reporting. Hellman, Patel, and Tsunogaya (2021) found significant differences in phrasing and expression of accounting concepts between German and UK financial disclosures produced in English. Roy and Mukherjee (2022) discovered that country-level culture, rather than economic factors, explained wide variations of corporate ESG disclosure practices across countries.

This thesis aims to answer the research question, “Do Swedish-based and UK-based companies differ in the way they display optimism in their written ESG disclosures?” by creating two corpora of ESG reports from Swedish (published in English) and UK-based companies.

2. Literature review

This section sets out to describe and summarize relevant academic papers for the theoretical framework used in this thesis. It will start with a primer on the field of corpus analysis, semantics, and text mining. Thereafter it will subsequently outline and describe the DICTION framework and will conclude with a short summary of previous research.

2.1. Corpus analysis, semantics, and text mining

This section sets out to build the theoretical linguistic framework the paper builds upon, for a more practical description please see the DICTION framework section as well as the method chapter.

2.1.1. Corpus analysis

Researchers Vaughan and O'Keefe (2015) describe corpus analysis as linguistics involving the use of computers to search and analyse data while Finegan (2008) describes it as a method for compiling collections of texts. A third, and maybe more comprehensive way of describing the field comes from Lindquist and Levin (2018, p.1) who write "Corpus linguistics is thus a methodology, comprising a large number of related methods which can be used by scholars of many different theoretical leanings". McCarthy and O'Keeffe (2010) observe that the field underwent significant changes and grew into its modern form with the introduction of enhanced storage capabilities, making the size of an individual corpus effectively endless. The methodology builds on generating a set of data, the corpora, which is then analysed to yield insights on the patterns of language use.

2.1.1.1. The corpus

Finnegan (2008) states that in a corpus analysis a 'corpus' is created by the collected written or transcribed spoken language. According to Yakute (2022), the corpus is the main data that a corpus linguist uses to investigate a specific area of a particular language, and it is not uncommon to use several corpora and compare these. Depending on the purpose of the research, different corpora can be developed. The focus in this thesis will be on two types of corpora, specialized and comparable corpora, as defined by Yakute (2022). A specialized corpus is a collection of texts from a specific field or genre (e.g., newspaper articles, lectures, essays, etc.) and comparable corpora is two or more sets of corpora from different languages or produced by similar but distinctly different authors / speakers (e.g., French, and English newspaper articles) (ibid).

2.1.2. Semantics

Semantics is described by Finegan as "the study of the systematic ways in which languages structure meaning, especially in words and sentences*"* (Finegan, 2008, p.546). Researchers Smith, Florence, and Maria (2018) elaborate on and define it as the study of meaning, as inherent at the levels of words, phrases, sentences, and larger units of discourse. Given the focus of the thesis on comparing the Optimism variable in different corpora through semantic analysis, it is important to establish a clear understanding of the field of semantics. As Finegan and researchers Smith, Florence, and Maria have described, semantics is the study of how language structures meaning, specifically at the levels of words, phrases, sentences, and larger units of discourse. With this foundation, the literature section below will delve deeper into the field of semantics, providing a framework for the analysis of individual words and their meaning in the context of the study's research question.

2.1.2.1. Lexical semantics: The meaning of individual words

Saeed (2016) defines lexical semantics as the study of the meaning of words and how they relate to each other within a language, while according to Finegan (2008), lexical semantics focuses on the lexical items, the actual written words, making up a specific language, and their meaning. Lexical semantics is concerned with investigating the meanings of individual words, as well as the ways in which words are combined to create more complex meanings. Saeed (2016) postulates that the meaning of a word is not fixed or universal, but rather it is shaped by its use in different contexts and by the conventions of the language in which it is used. Therefore, lexical semantics is concerned with understanding how words acquire meaning through their use in context (further described in the denotation and connotation section), and how their meanings may vary across different contexts and cultures.

2.1.2.2. Lexical semantics: Denotation and connotation

According to Finnegan (2008), the denotation of an expression, often referred to as its linguistic meaning, stands in contrast to its connotation, which encompasses social and affective meanings. Both denotation and connotation will be explored in the context of the DICTION framework later in this section, but the text below introduces the concepts.

Denotation: Finnegan (2008) describes the concept denotation as the primary, literal, or explicit meaning of a word. It is the objective definition of a word, independent of any emotional or cultural associations that may be attached to it. Denotation is key for understanding how words function within a language and how they can be used accurately in different contexts. For example, consider the word 'House'. Its denotative meaning is a building for human habitation, usually consisting of rooms and facilities for people to live in. This meaning remains consistent across various contexts and is universally understood by speakers of the language. By recognizing the denotative meaning of words, we can ensure that our communication is clear and effective, minimizing potential misunderstandings.

Connotation: Connotation, on the other hand, is explained by Finnegan (2008) as the implicit, emotional, or associative meanings that a word carries beyond its literal definition. These meanings are shaped by cultural, historical, and personal experiences and can vary among individuals or communities. Returning to the example of "house," while its denotative meaning remains constant, its connotative meaning can differ depending on the context and individual associations.

2.1.2.3. Lexical typology: How languages differ in the ways in which they encode meaning

According to Saeed (2016), lexical typology refers to the examination of linguistic variations in vocabulary across different languages. It explores how languages encode meaning through their vocabulary (Ibid). In this thesis, although both corpora analysed are in English, the Swedish-based corpus was initially written in Swedish and then translated into English. Therefore, some remnants of Swedish word choices may persist to ensure the fidelity of the translated text. While lexical typology is not the field under investigation it might be an explanatory factor to differences in findings and is therefore included.

2.1.3. Text mining

This section will introduce text mining as a linguistic tool to discern the level of optimism, quantified as the Optimism variable, in the analysed corpora. Hotho, Nürnberger and Paass postulate that: "text mining aims at disclosing the concealed information by means of methods which on the one hand are able to cope with the large number of words and structures in natural language and conversely, allowing the handling of vagueness, uncertainty, and fuzziness (2005, p.19)". From the context of corpus analysis, the text mining is the tool applied to the corpus, e.g., the method of which data to be analysed is extracted from a specific corpus. Hickman et al. (2022) claim that text mining provides a new venue for capitalizing on the extensive corpus of text data that organizations, their customers and their employees generate.

2.1.3.1. Closed Vocabulary Text Mining

This thesis will use what Hickman et al (2022) describe as Closed Vocabulary Text Mining. This is a method in which words and phrases are counted and compared with a generated dictionary (Ibid). These dictionaries, or word banks, are generally built with a specific purpose in mind and aim to capture a specific construct. They can range from very simplistic, e.g., counting how many times one specific word is used, to more complex, e.g., assigning scores based on a multiplicity of words used – and sometimes words not used. The output is often further processed in statistical modelling such as regressions and analysis of variance. An important prerequisite for Closed Vocabulary Text Mining to be effective is that the corpus analysed is not pre-processed. This means that the researcher does not change or alter the content. If such actions are undertaken, the integrity of the research is jeopardized as the original usage of lexemes are lost. In contrast to the closed vocabulary approach, there is also Open Vocabulary Text Mining, in which the researcher does not have a dictionary with pre-defined vocabulary (but rather uses the method to develop one).

2.2. The DICTION framework

According to Hart (2001), DICTION is a software program created by James W. Pennebaker, and his colleagues at the University of Texas. It is designed to analyse written text for a range of linguistic and psychological features (Ibid). Hickman et al (2022) define the program as a closed vocabulary dictionary-based approach to text analysis. In DICTION each word in a text is assigned a score based on its frequency and association with various linguistic and psychological categories (Ibid). DICTION is a tool built around 31 predefined dictionaries consisting of c 10 000 words in total. Each word has a denotation and is either positively, or negatively, associated with one of five scales, the so called ‘master variables’, Activity, Optimism, Certainty, Realism and Commonality (Ibid). The software will create a profile for each analysed text by counting the frequency of words. It balances the positive denotation of a word (e.g., strong) with the negative denotation (e.g., weak) and comparing these against the specific master variable dictionaries (Table 1 contains additional details on the Optimism master variable). The word lists are predominantly made up by the word classes nouns, adjectives, verbs, and adverbs, but also to a certain extent other word classes such as determiners (numerical terms) and pronouns (overt use of I, mine, and me). By comparing a text against the pre-defined dictionaries, the software scores each text on its level of each of the master variables. The DICTION framework normalizes its output to ensure comparability between corpora of different length and volume. This is done by scoring every section of 500 words, once the text has been divided into sections of 500 words these will be assigned a score, and then the text’s score will be the average of the number of scores. This yields a score which can be compared to other texts, regardless of their length or volume.

2.2.1. Optimism

This master variable reflects the degree of positive impact in the language used. It captures the extent to which the text is characterized by optimistic language, positive emotion words, and a generally hopeful tone. Optimism is associated with six specific dictionaries: three positively associated with the trait, and three negatively associated.

Table 1. The six dictionaries associated with the Optimism master variable.

Dictionary	Description	Examples (Hart, 2001) *
+ **Praise**	Words associated with positive evaluation, praise, and approval. It aims to capture expressions of positive sentiment or approval towards oneself or others.	Social qualities (dear, delightful, witty), physical qualities (mighty, handsome, beautiful), intellectual qualities (shrewd, bright, vigilant, reasonable), entrepreneurial qualities (successful, conscientious, renowned), and moral qualities (faithful, good, noble). All terms in this dictionary are adjectives.
+ **Satisfaction**	Words associated with contentment, fulfilment, and satisfaction. It aims to capture expressions of satisfaction or positive emotions related to achievement, success, or comfort.	Terms associated with positive affective states (cheerful, passionate, happiness), with moments of undiminished joy (thanks, smile, welcome) and pleasurable diversion (excited, fun, lucky), or with moments of triumph (celebrating, pride, auspicious). Also included are words of nurturance: healing, encourage, secure, relieved.
+ **Inspiration**	words associated with creativity, imagination, and inspiration. It aims to capture expressions of motivation, inspiration, or positive emotions related to imagination or creativity	Abstract virtues deserving of universal respect. Most of the terms in this dictionary are nouns isolating desirable moral qualities (faith, honesty, self-sacrifice, virtue) as well as attractive personal qualities (courage, dedication, wisdom, mercy). Social and political ideals are also included: patriotism, success, education, justice.
- **Blame**	words associated with blame, responsibility, and fault-finding. It aims to capture expressions of negative sentiment or criticism towards oneself or others.	Terms designating social inappropriateness (mean, naive, sloppy, stupid) as well as downright evil (fascist, blood-thirsty, repugnant, malicious) compose this dictionary. In addition, adjectives describing unfortunate circumstances (bankrupt, rash, morbid, embarrassing) or unplanned vicissitudes (weary, nervous, painful, detrimental) are included. The dictionary also contains outright denigrations: cruel, illegitimate, offensive, miserly.
- **Hardship**	contains words associated with hardship, adversity, and struggle. It aims to capture expressions of negative emotions related to difficulty, challenge, or adversity.	This dictionary contains natural disasters (earthquake, starvation, tornado, pollution), hostile actions (killers, bankruptcy, enemies, vices) and censurable human behaviour (infidelity, despots, betrayal). It also includes unsavoury political outcomes (injustice, slavery, exploitation, rebellion) as well as normal human fears (grief, unemployment, died, apprehension) and in capacities (error, cop-outs, weakness).

- **Denial**	words associated with denial, rejection, and disbelief. It aims to capture expressions of negative sentiment or disbelief towards oneself or others	A dictionary consisting of standard negative contractions (aren't, shouldn't, don't), negative functions words (nor, not, nay), and terms designating null sets (nothing, nobody, none).

* *The examples are collected from the DICTION framework as presented by Hart (2001) to keep the exact meaning as close as possible.*

2.2.2. Limitations

One risk of applying an approach based on closed vocabulary is that the ambiguous words might be misinterpreted. For example, the word 'Interesting' could be positively or negatively coded depending on the context in which it is used. Such words may require additional analysis or interpretation to determine their true coding in the context of a particular text or speech. Similarly, a closed vocabulary approach focuses on denotation and may not capture the connotation of a specific word, potentially skewing the output and making any analysis based on it potentially skewed.

3. Methods

3.1. Theoretical foundation

The approach selected for this study is a blend of qualitative discourse analysis and a quantitative methodology. The qualitative discourse analysis is provided by the creators of the DICTION tool. The assignment of value and denotation of each of the 10,000 words used in the analysis is a qualitative assessment. This thesis approaches the method from a purely quantitative perspective as the author does not influence the DICTION dictionary. Quantitative research is described by Bell, Bryman, and Harley (2019) as “approaches that attempt to measure and / or count social phenomena and the relationship between them”. The same authors describe discourse analysis as: “an approach to analysis of talk and other forms of language that emphasizes the way in which versions of reality are accomplished through language” (ibid). Bell et al (2019) noted that the division between purely qualitative and quantitative research designs becomes ambiguous in practice. This study integrates qualitative assessments - analysing the meaning of words in Hart's DICTION-framework (Hart, 2001) - and quantitative measures - collecting and comparing various language expressions- the method can be applied on a great number of texts easily to try to answer the research question. This research builds on two corpora, one for Swedish-based companies and one for UK-based companies, both have been developed for the purpose of this thesis. This approach sets out to compare the level of optimism variable used in language of Swedish and UK-based companies when disclosing their ESG reporting.

3.2. Methods of data collection

To do this, 21 Swedish and 21 UK-based ESG disclosures has been used. To increase the level of comparability these have been selected using several criteria to ensure that the two populations share as many characteristics as possible. The criteria are:

1. **Publicly traded companies:** All companies in the thesis will be publicly traded, this is due to the level of scrutiny and the legal regulation of disclosure requirements.
2. **Same industry:** To ensure that the companies in the sample face the same challenges one industry sector has been selected, the ESG challenges for an industrial manufacturer would for instance vastly differ from the ones of a professional service company. Since the UK and Sweden have several listed financial institutions, this industry is chosen out of convenience purposes. The disclosure legislation has been in place since 2014 (NFRD (2014/95/EU).

3. **Time period**: To ensure the broadest possible data set, a longitudinal approach has been adopted. Therefore, all companies in scope of the study's ESG disclosure have been collected from 2019 and onward, to create as wide a sample as possible.

ESG disclosures were collected from the annual reports found on the 'Investor' section of the listed companies' websites. The annual reports are located under the Financial Reporting section. The ESG disclosure is typically labelled as "ESG Disclosure", "ESG Report", or "Sustainability Report", for further information please see Appendix I.

Each text is either part of the Swedish-based or the UK-based corpus and will be categorized as such. These categorisations will be used to compare the 'level of optimism variable' in the ESG disclosures of the countries. Table 2 outlines the companies used in the study. As described previously both Swedish and UK-based companies in the same industry have been chosen. These have been ordered in size (sized as measured by market capitalization). Each company will have three ESG disclosures, issued in 2020, 2021, and 2022 respectively. However, for firms that have not issued their 2022 ESG disclosure at the time of writing, their 2019 disclosure will be included instead. With seven listed companies per country, this yields a total of 21 ESG disclosures per country.

3.2.1. The UK corpus

The UK sample comprises seven companies, totalling 21 annual ESG disclosures for the relevant time periods. These companies operate in both the banking and insurance sectors.

Table 2. UK-based companies within the scope of the thesis.

Company	#Characters analysed	#Words analysed	#Average word size	#Unique words	#Numerical terms
Aviva	355 486	52 410	5.38	24 059	165.86
Barclays	927 442	136 669	5.51	63 131	74.92
HSBC	712 135	107 820	5.29	49 686	124.04
Legal & General	606 613	91 344	5.32	42 197	61.49
Lloyd's bank	858 061	129 281	5.33	60 079	328.78
Prudential	1088 123	159 444	5.49	73 789	666.31
RBS	1100 203	164 999	5.39	73 308	313.24
TOTAL	5 648 063	841 967	5.39	386 249	1 734.64

#Characters analysed: *Number of character's analysed consist of characters in the text excluding spaces, punctuation marks, or other non-word characters.*
#Words analysed: *The number of words analysed refers to the count of individual words analysed in a specific text.*
#Average word size: *Average word size is calculated by summing the length of words and dividing by three, for each company.*
#Unique words*: For average word size per domicile, the average word size per company is added and divided by seven.*
#Numerical terms: *The number of numerical terms refers to the count of words referring to numeric values (such as 5 or five) in each analysed text.*

3.2.2. The Swedish corpus

The Swedish sample consists of seven companies with a total of 21 annual ESG disclosures in the relevant time periods. The companies operate in the banking, insurance, and debt collection sectors.

Table 3. Swedish-based companies within the scope of the thesis.

Company	#Characters analysed	#Words analysed	#Average word size	#Unique words	#Numerical terms
Collector	115 806	17 196	5.48	8 116	153.61
Folksam	564 072	86 468	5.25	38 903	120.63
Handelsbanken	728 963	108 346	5.42	50 626	41.13
If	627 722	93 125	5.23	43 026	218.94
Nordea	319 995	48 754	5.24	21 474	239.1
SEB	343 280	51 062	5.44	23 706	70.22
Swedbank	624 801	95 813	5.24	40 548	325.55
TOTAL	3 324 639	500 764	5.33	226 399	1 169.18

#Characters analysed: *Number of character's analysed consist of characters in the text excluding spaces, punctuation marks, or other non-word characters.*
#Words analysed: *The number of words analysed refers to the count of individual words analysed in a specific text.*
#Average word size: *Average word size is calculated by summing the length of words and dividing by three, for each company.*
#Unique words*: For average word size per domicile, the average word size per company is added and divided by seven.*
#Numerical terms: *The number of numerical terms refers to the count of words referring to numeric values (such as 5 or five) in each analysed text.*

3.2.3. Comparing the two corpora

Figure 1, below, outlines the difference between the two corpora.

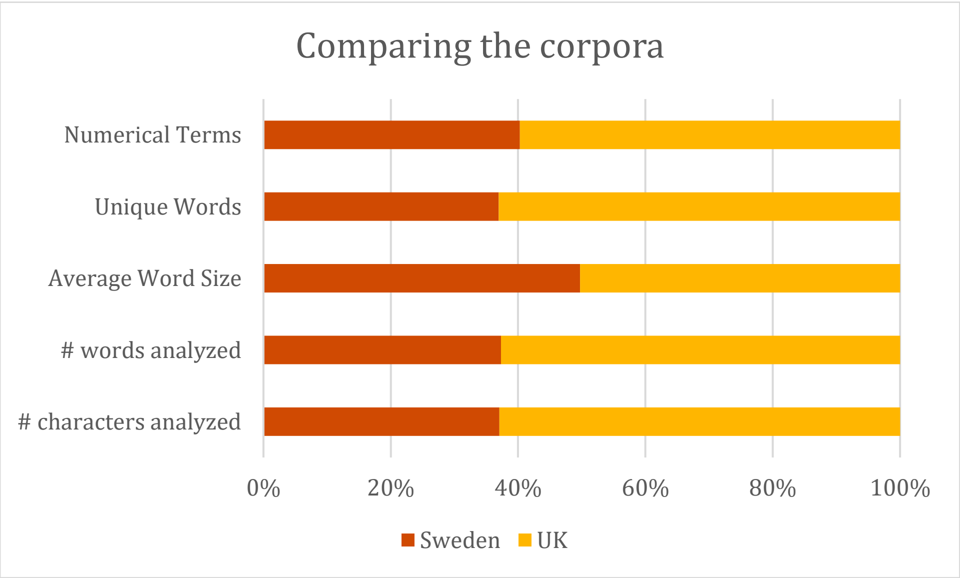


Figure 1. Comparative statistics on the two corpora

The UK corpus is lengthier compared to the Swedish one comprising 63 % of the total characters analysed and 63 % of the total words analysed. The average word size of the two corpora is approximately the same (5.33 characters per word for the Swedish and 5.39 for the UK). The UK corpus uses a significantly greater number of unique words per text, but if these are normalized with the total number of words the number of unique words is similar for both texts, with the UK corpus making up 69 % of the total number of unique words.

Table 4. Normalization and comparison between unique words and numerical terms.

	Normalization of unique words			Normalization numerical terms		
	Unique words	**Total Words**	**Normalized by 2**	**Numerical Terms**	**Total words**	**Normalized by 2**
SE	226 399	500 764	0.4500	1 169.18	500 764	0.0023
UK	386 249	841 967	0.4600	1 734.64	841 967	0.0020

3.3. Methods of data analysis

3.3.1. Quantifying the optimism variable

To quantify the Optimism level per text, DICTION compares each text against the six sub-dictionaries, a set of word lists with value coded denotations per word. This approach is further described in the literature review section summarizing the DICTION framework. The process of scoring each text proceeds as follows:

1. **Calculate number of words per sub-dictionary:** Each time a specific word occurs in the text associated with a sub-dictionary this word is counted. If the word occurs more than once it is counted as many times as it is mentioned.
2. **Conversion to z-values for comparability:** Each sum from the sub-dictionary is converted to a z-value (a comparable value based on the normal distribution of occurrence in the sub-dictionaries).
3. **Summing the positively associated sub-dictionaries and summing the negatively associated sub-dictionaries:** Once the z-values are compiled these can be summarized into the positively associated sub-dictionaries and the negatively associated sub-dictionaries.
4. **Calculating the penultimate score:** The negatively associated z-values are subtracted from the positively associated ones to obtain the penultimate score.
5. **Add 50:** A value of 50 is added to the score to avoid negative values (this is only a corrective measure to make the figures easier to work with).

An alternative way of describing the process as a formula is:

$$\boldsymbol{Optimism} = (Praise + Satisfaction + Inspiration) - (Blame + Hardship + Denial) + 50$$

As a final step the DICTION software makes a slight adjustment to the score using DICTION's normative databank to provide a statistical correction.

3.3.2. The T-test: Testing if the means of the corpora differ

A two sample T-test will be applied to the two population profiles to test for significant differences in the ESG disclosure level of optimism between Swedish and UK companies. This test, as described by Bell et al, is used to determine if there is a statistically significant difference between the means of two independent groups (2019).

The two-sample T-test is, according to Bell et al, applied as follows:

- **Null and alternative hypotheses:** The null hypothesis states that there is no significant difference between the means of the two groups, while the alternative hypothesis states that there is a significant difference.
- **Test statistic calculation:** The test statistic for the two-sample T-test is calculated by taking the difference between the means of the two groups and dividing it by the standard error of the difference.
- **Degrees of freedom:** The degrees of freedom are calculated by subtracting one from the sample sizes of each group, and then taking the smaller of the two values.
- **The p-value calculation:** The p-value is calculated using the t-distribution with the degrees of freedom and the test statistic.
- **Comparison between the p-value to the level of significance:** If the p-value is less than the level of significance (usually set at 0.05), then we reject the null hypothesis and conclude that there is a significant difference between the means of the two groups. Otherwise, we fail to reject the null hypothesis and conclude that there is no significant difference.

Below outlines how the two-sided T-test is applied in the context of this thesis:

- **Null and alternative hypotheses:**
  - $H_0: \mu_1 = \mu_2$, or all means are equal = there is no difference in level of optimism between the UK-based and the Swedish companies' ESG disclosure
  - $H_1: \mu_1 \neq \mu_2$ or at least one mean differs between the groups = there is difference in level of optimism between the UK-based and the Swedish companies' ESG disclosure
- **Test statistic calculation:** Please see subsequent sections.
- **Degrees of freedom:** Please see subsequent sections.
- **The p-value calculation:** Please see subsequent sections.

- **Comparison between the p-value to the level of significance:** Confidence interval set to 95 %

Bell et al (2019) state that for the T-test to yield accurate results, a set of assumptions must hold true. These include:

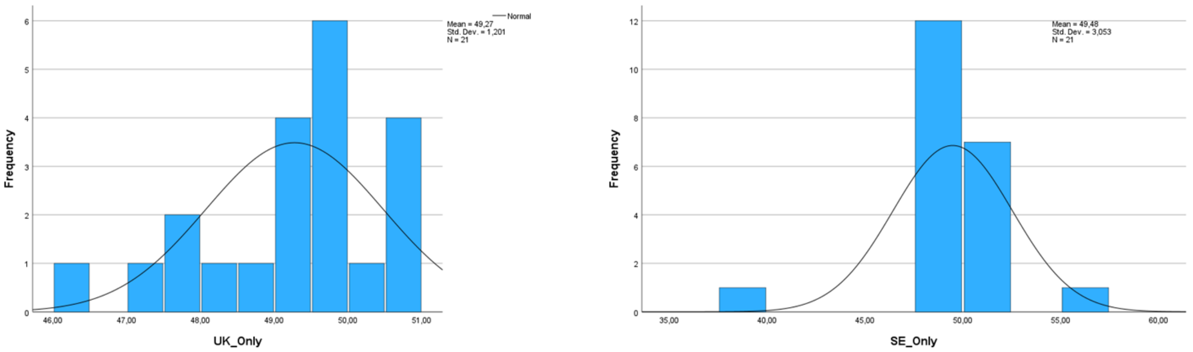
- **Independence:** The observations in each group must be independent of each other.
- **Normality:** The data within each group must be normally distributed.
- **Homogeneity of variance:** The variance of the data within each group must be equal or approximately equal.
- **Scale of measurement:** The data must be measured on a continuous or ordinal scale.

It is important to check these assumptions before conducting a T-test to ensure that the results are valid and reliable. Violations of these assumptions can lead to inaccurate or misleading results, and alternative statistical tests may be necessary.

3.3.3. Verifying whether the assumptions hold true

To verify the assumptions, the following steps have been undertaken.

- **Independence:** As each observation is unique, and there is no interrelationship between the texts used, these criteria are assumed to be true.
- **Normality:** Given the size of the corpora (21 observations in each), this is likely on the lower end of the spectrum. However, to validate this assumption, two histograms were produced.



UK Normal distribution **SE Normal distribution**

Figure 2. The distribution of UK and Swedish corpora, plotted as histograms using the statistical tool SPSS

- **Homogeneity of variance:** Levene's test, a statistical procedure, can assess homogeneity of variance. This test is used to determine if the variances of the data within each group are equal or approximately equal. Upon performing the T-test, both the F-value and significance value suggested that there was no significant difference in the variances between the two groups (significance greater than 0.05).

Table 5. The output from SPSS for Levene's test

	F-value	Significance
Equal variances assumed	0.747	0.393

- **Scale of Measurement:** Both corpora utilize the same continuous scale, namely optimism, as measured by the DICTION framework.

In summary, it appears that these assumptions are valid, hence, the T-test can be performed.

4. Results

4.1. Comparing the two corpora

4.1.1. Scoring by sub-dictionary

The first step of creating the optimism variable score is to develop the sub-dictionary scoring. Tables 6 and 7 outline the per company per year scoring after comparing the ESG disclosures with the DICTION sub-dictionary word lists and converting these to z-values. The z-values in Table 6 illustrates the differences in ESG disclosures usage across each of the sub-dictionaries. In a subsequent step, these z-values will be aggregated to form the complete optimism variable, as described in section three.

Table 6. DICTION output and scoring per ESG disclosure per year.

Company	D	Y	Praise	Satisfa.	Insp.	Blame	Hardship	Denial	Factor	Adjust.
Collector	**SE**	**19**	4.30	0.43	6.24	0.60	6.19	1.55	50	-18.97
Collector	**SE**	**20**	5.07	3.00	2.99	0.19	5.53	2.48	50	-18.36
Collector	**SE**	**21**	0.77	0.26	1.09	0.59	2.45	0.41	50	-6.29
Folksam	**SE**	**19**	0.03	0.89	6.96	0	3.02	0.08	50	-10.22
Folksam	**SE**	**20**	3.89	0.38	4.08	0.25	1.27	5.17	50	-14.98
Folksam	**SE**	**21**	2.70	0.38	2.57	0.25	1.40	5.98	50	-14.12
Handelsbanken	**SE**	**20**	1.76	0.32	3.56	1.20	2.76	0.34	50	-10.18
Handelsbanken	**SE**	**21**	0.21	0	3.47	0	7.49	0.77	50	-13.32
Handelsbanken	**SE**	**22**	2.28	0.24	3.54	0.39	0.79	0.08	50	-6.56
If	**SE**	**19**	8.87	7.97	2.56	0.88	0.73	0.16	50	-16.17
If	**SE**	**20**	2.38	1.58	6.65	0.38	1.21	3.03	50	-14.15
If	**SE**	**21**	2.70	1.68	7.55	0.88	1.50	1.44	50	-14.25
Nordea	**SE**	**20**	0.75	0.63	2.74	0.14	2.85	1.14	50	-8.56
Nordea	**SE**	**21**	0.57	0.39	1.24	0.06	2.33	46.93	50	-63.90
Nordea	**SE**	**22**	0.34	0.28	1.94	0.18	0.87	3.87	50	-8.49
SEB	**SE**	**20**	0.51	0.02	6.13	0.03	8.63	2.83	50	-19.67
SEB	**SE**	**21**	1.25	0.05	4.91	0.16	6.05	0.91	50	-13.89
SEB	**SE**	**22**	1.37	0.38	3.36	0.51	3.24	2.34	50	-11.86
Swedbank	**SE**	**20**	2.29	0.41	4.62	0.31	6.55	1.88	50	-16.67
Swedbank	**SE**	**21**	1.14	0.30	4.26	0.22	5.82	0.10	50	-12.25
Swedbank	**SE**	**22**	0.88	0.20	4.29	0.16	5.89	0.95	50	-13.09
Aviva	**UK**	**20**	0.91	0.39	5.82	0.25	6.38	3.44	50	-18.30
Aviva	**UK**	**21**	5.05	1.49	5.27	0.32	4.13	3.55	50	-19.01
Aviva	**UK**	**22**	6.69	1.54	4.35	0.02	3.48	5.26	50	-20.38
Barclays	**UK**	**20**	2.50	3.83	2.53	0.22	6.34	3.52	50	-19.04
Barclays	**UK**	**21**	6.81	1.68	3.70	4.46	6.10	0.06	50	-23.30
Barclays	**UK**	**22**	0.64	0.02	1.90	0.23	9.15	6.56	50	-22.03
HSBC	**UK**	**20**	4.09	2.93	4.78	0.34	6.34	2.90	50	-20.76
HSBC	**UK**	**21**	2.95	1.32	3.14	0.25	7.21	2.31	50	-17.81
HSBC	**UK**	**22**	1.69	3.64	5.38	0.17	11.52	0.41	50	-22.92

Legal&General	**UK**	**20**	1.45	0.73	4.42	0	2.46	0.45	50	-8.89
Legal&General	**UK**	**21**	1.01	0.14	3.43	0.51	5.73	7.34	50	-20.81
Legal&General	**UK**	**22**	1.27	1.36	2.92	0.49	5.80	7.43	50	-21.55
Lloyd's bank	**UK**	**20**	0.62	0.06	3.00	0.02	0.97	3.10	50	-8.30
Lloyd's bank	**UK**	**21**	1.16	1.42	3.23	0.10	3.38	2.34	50	-11.84
Lloyd's bank	**UK**	**22**	0.52	1.10	4.94	0.77	7.39	3.81	50	-20.25
Prudential	**UK**	**20**	0.83	0.02	0.94	0.03	0.11	0.26	50	-2.19
Prudential	**UK**	**21**	0.18	0.20	0.41	0.06	0.68	0.12	50	-1.99
Prudential	**UK**	**22**	0.27	0.04	0.21	0.03	0.19	0.40	50	-1.50
RBS	**UK**	**20**	0.79	0.01	1.47	0.27	4.14	0.55	50	-8.20
RBS	**UK**	**21**	0.83	0.01	1.80	0.12	4.03	0.64	50	-8.25
RBS	**UK**	**22**	1.43	0.16	3.55	0.22	3.65	9.42	50	-20.87

D = Domicile, Y = Year of ESG Disclosure. The sub-dictionaries names have been abbreviated in the table above, they are in order: Praise, Satisfaction, Inspiration, Blame, Hardship and Denial. In addition to the sub-dictionaries the above table also includes (i) the factor, 50, which is used to turn the figures positive and (ii) the adjustment factor as applied by DICTION to statistically correct the scoring.

Table 7 shows an aggregated average z-value result per sub-dictionary, per corpora. These z-values will be used to examine if the two corpora statistically differ in any of the components of the optimism variable. E.g., it might be that they do not differ on the full optimism variable, but that for instance the praise sub-dictionary component differs.

Table 7. Average z-value per sub-dictionary.

	Positive			Negative		
Average	**Praise**	**Satisfaction**	**Inspiration**	**Blame**	**Hardship**	**Denial**
SE	2.10	0.94	4.04	0.35	3.65	3.93
UK	2.04	1.09	3.07	0.43	4.64	3.02
Total	**2.07**	**1.01**	**3.55**	**0.39**	**4.14**	**3.47**

4.1.2. The optimism variable scores

Transitioning from the sub-dictionary level outlined in previous section, Table 8 provides an overview of the optimism variable, per ESG disclosure. This is done by combining the sub-dictionary z-values, as described in section three. This will be the main component of the analysis, as the means of the two corpora and their standard deviation will be used to assess if there is a statistically significant difference between the two populations' optimism variables. The below output has been produced by the DICTION software:

Table 8. Optimism variable score per ESG disclosure

SE company	Year	Optimism	UK company	Year	Optimism
Collector	2019	50.34	**Aviva**	2020	48.89
Collector	2020	50.90	**Aviva**	2021	50.80
Collector	2021	49.28	**Aviva**	2022	50.96
Folksam	2019	50.76	**Barclays**	2020	49.90
Folksam	2020	50.06	**Barclays**	2021	49.51
Folksam	2021	49.16	**Barclays**	2022	46.47
Handelsbanken	2020	49.76	**HSBC**	2020	50.62
Handelsbanken	2021	48.62	**HSBC**	2021	49.37
Handelsbanken	2022	50.76	**HSBC**	2022	49.89
If	2019	55.00	**Legal & General Group Plc**	2020	50.62
If	2020	51.08	**Legal & General Group Plc**	2021	47.35
If	2021	51.50	**Legal & General Group Plc**	2022	47.72
Nordea	2020	49.69	**Lloyd's bank**	2020	49.47
Nordea	2021	37.62	**Lloyd's bank**	2021	49.79
Nordea	2022	48.99	**Lloyd's bank**	2022	48.28
SEB	2020	48.48	**Prudential**	2020	50.00
SEB	2021	49.44	**Prudential**	2021	49.66
SEB	2022	49.34	**Prudential**	2022	49.64
Swedbank	2020	49.39	**RBS**	2020	49.03
Swedbank	2021	49.59	**RBS**	2021	49.18
Swedbank	2022	49.28	**RBS**	2022	47.56

4.1.3. Group statistics

As a first step the groups are compared along several metrics. In Table 9, an extraction from the statistics software SPSS, the different columns have the following meaning:

- **N:** N denotes the number of observations in a population (in this case the number of texts analysed in each corpora).
- **Mean:** Mean is a statistic term for the average of a population, in this case the average in the Optimism variable, as measured by DICTION from the two corpora's DICTION output.

- **Std. Deviation:** Is the standard deviation in the distribution of optimism variable scores for each corpus. Standard deviation is a measure of how spread out a set of data is from the average or mean value. It indicates much the individual data points deviate from the average, with higher standard deviation indicating more variability or dispersion within the data.
- **Std. Error Mean:** The standard error of the mean is a measure of the variability of sample means when repeatedly sampling from the same population. It informs how much the means are likely to differ from the true population mean due to random sampling error. A smaller standard error of the mean indicates that the sample means are more precise and closer to the true population mean, while a larger standard error of the mean indicates that the sample means are less precise and more likely to be further from the true population mean due to random variation in the sampling process.

Table 9 provides a statistical overview of the two corpora; these metrics will be used when conducting the two-sided T-test which will be applied to assess if there is a statistical difference between the two populations' optimism variables. As seen in Table 9, the sample comprises 42 texts analysed. The two samples' means are very close (only differing 0.30 points on the Optimism variable scale). The standard deviation is significantly greater in the Swedish sample, indicating that the spread in the Optimism variable values is significantly greater. Finally, the standard error of the mean is very low, indicating that the sample means are precise.

Table 9. Statistics of the two corpora

Domicile	N	Mean	Std. Deviation	Std. Error Mean
SE	21	49.48	3.05340	0.66631
UK	21	49.27	1.20138	0.26216

4.1.4. The sub-dictionaries

The output of the sub-dictionaries has been mapped per year, per corpora and in total. Table 10 displays the development in z-values per year, for a graphical representation of the development, please refer to Appendix III. These figures (Figure 4 and Figure 5) thus indicate the development of the different subcomponents building up the Optimism variable, on a year-by-year basis. This is relevant since it might be that on aggregate, when combining all the sub-dictionary z-values, there is no difference between the two corpora. However, if analysed on a sub-dictionary level it might be that the corpora differ in for instance the z-value mean of the blame sub-dictionary.

Table 10. Per year, per corpora and total, sub-dictionary development

Positive					Negative				
Praise	**2019**	**2020**	**2021**	**2022**	**Blame**	**2019**	**2020**	**2021**	**2022**
SE	4.40	2.37	1.33	1.22	**SE**	0.49	0.35	0.30	0.31
UK	N.a.	1.59	2.57	1.79	**UK**	N.a.	0.16	0.83	0.27
Total	4.4	1.98	1.95	1.58	**Total**	0.49	0.25	0.57	0.28
Satisfaction	**2019**	**2020**	**2021**	**2022**	**Hardship**	**2019**	**2020**	**2021**	**2022**
SE	3.09	0.90	0.43	0.27	**SE**	3.31	4.11	3.86	2.69
UK	N.a.	1.13	0.89	1.12	**UK**	N.a.	3.82	4.46	5.88
Total	3.09	1.02	0.66	0.81	**Total**	3.31	3.96	4.16	4.72
Inspiration	**2019**	**2020**	**2021**	**2022**	**Denial**	**2019**	**2020**	**2021**	**2022**
SE	5.25	4.39	3.58	3.28	**SE**	0.59	2.41	8.07	1.81
UK	N.a.	3.28	2.99	3.32	**UK**	N.a.	2.03	2.33	4.75
Total	5.25	3.83	3.29	3.30	**Total**	0.59	2.22	5.20	3.68

4.1.5. Comparing the sub-dictionaries

The two corpora have been tested using a two-sided T-test, comparing each of the means of the six sub-dictionaries, to determine if there are any differences in the countries ESG disclosures. The full SPSS output is summarized in Table 11 and included as full SPSS output in Appendix II. As seen in Table 11, there is no statistically significant difference between the mean of the corpora, as the p-value is not greater than 0.05.

Table 11. Two-sided T-test output per sub-dictionary

Sub-dictionary	T-value	DF	Two-sided p-value
Praise	0.179	40	0.859
Satisfaction	-0.237	40	0.814
Inspiration	1.557	40	0.127
Blame	-0.328	40	0.745
Hardship	-1.277	40	0.209
Denial	0.391	40	0.698

4.2. The combined Optimism variable

The development of the level of Optimism variable per year, per corpora, has been mapped to analyse if there are any trends in the underlying data.

Table 12. Comparison of level of Optimism variable per year per corpora

Average level of Optimism per year	2019	2020	2021	2022
SE	52.03	49.91	47.88	49.59
UK	N.a.	49.79	49.38	48.64
Total	52.03	49.84	48.63	48.99

4.2.1. Applying the two-sided T-test to the combined Optimism variable

As outlined in the previous section a two-sided T-test with a confidence interval of 95 % has been applied when comparing the means of the two populations. The combined mean of the total UK and SE sample has been used, making the full sample 21 firm year observations per corpora. This has been done by using SPSS, a tool for statistical analysis. Figure 3, a screen capture from SPSS' output viewer, indicate that the null hypothesis cannot be rejected. This will be further elaborated upon in the following sections.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
Optimism	Equal variances assumed	,747	,393	,288	40	,387	,775	,20619	,71603	-1,24095	1,65334
	Equal variances not assumed			,288	26,047	,388	,776	,20619	,71603	-1,26549	1,67788

Figure 3. Output from SPSS statistic tools when comparing the two populations means using a two-sided T-test

The first section details the Levene's test for equality of variances to evaluate the assumption on homogeneity. This is further described in section three, outlining the method. The second section, T-test for Equality of Means, provide the results of the T-test. The two-sided p-value indicates that there was no significant difference between the means of the two groups, as it breaches the confidentiality threshold of 0.05.

5. Discussion

5.1. Initial reflections

UK ESG disclosures, making up 63 % of all analysed words, are considerably longer than their Swedish counterparts. However, upon normalizing for the unique number of words used and numerical terms used in the texts, both corpora exhibit similar characteristics (see Table 4). Thus, it appears that both text corpora share a similar style, with the UK ESG disclosures placing a particular emphasis on length.

5.2. The sub-dictionaries

While there is no statistically significant difference between the two corpora in any of the six sub-dictionaries, when analysing the means, there are some interesting trends in the year-by-year developments. The usage of words associated with blame in the UK ESG disclosures increased from 0.6 in 2020 to 0.83 in 2021 to return to 0.27 in 2022. This is an increase of 4.15 in one year with the Swedish equivalent showing very low change in the same period. Similarly, the usage of words associated with Denial in the SE ESG disclosure increased from 2.41 in 2020 to 8.07 in 2021 to return to 1.81 in 2022 – a pattern the UK ESG disclosure did not showcase. The driving factors behind this can only be speculated at this point. However, it is worth noting that many financial service companies, such as banks offering credit cards and unsecured debt, struggled with reduced volumes in 2021 due to the global pandemic. If this is indeed the reason for the difference in the sub-dictionaries associated with Optimism variable between the years, it would indicate that Swedish-based financial services companies and UK-based equivalent have different communication strategies when framing the situation.

5.3. Testing the hypothesis

The T-test performed on the 42 ESG disclosures indicate that the null hypothesis cannot be rejected. This indicates that there, according to this test and this set of data, is not statistically (applying a threshold of 95 %) significant difference between the means of the two populations. Expressed clearly – there does not seem to be any difference in the level of Optimism variable used in UK-based companies and Swedish-based companies in their ESG disclosure, when publishing their annual English ESG reports.

5.4. Comparison with previous research

Albitar, Abdoush and Hussainey (2021) identified a difference in how UK-based and German based firms described accounting concepts. This finding might not pair with the finding that there is no difference in how UK-based and Swedish-based ESG disclosures. It might be that accounting concepts, being a long and established tradition hailing from the 14th century Italy, have developed, and branched off with local interpretations, while ESG disclosure being a novel concept invented and implemented at the same time in a globally connected economy are not. If provided with time to grow in isolation perhaps similar findings to those of Albitar, Abdoush and Hussainey could develop also in the ESG realm. Nesova and Larkin found that ESG reporting has higher levels of ‘Optimism’ than other corporate disclosure texts (2022). This however is not contrary to the findings in this thesis, as this material only deals with comparisons between different ESG disclosures – it might be that if comparing these ESG disclosure corpora with a corpus made up by other corporate disclosure, this body of text also exhibits increased levels of Optimism variable.

5.5. Limitations of approach

Although the T-test indicates that the null hypothesis cannot be rejected, it is important to note the following limitations to the applied approach:

1. Sample size: The sample used consists of 21 ESG disclosures from each country. To increase the level of robustness of the findings, this sample might be augmented with additional ESG disclosures to create a higher degree of reliance.
2. Sample industry: The thesis focuses on financial service companies based in the two countries. Expanding the sample to include other industries might further strengthen the findings.
3. Author of the ESG disclosures: Although this thesis does not comment on who wrote the specific companies’ ESG disclosure, it is important to note that this might influence the results. For instance, if Swedish-based companies all used UK-based PR firms with UK native authors accustomed to writing ESG disclosures on behalf of UK-based companies, this might skew the result. Similarly, if UK-based firms used non-UK-based PR firms for the same purpose (e.g., US based), this might also skew the findings.

6. Conclusion

6.1. The null hypothesis cannot be rejected

The text-analysis, and the subsequent T-test, indicate that there is no difference, in the level of Optimism variable, between the two sets of corpora. Thus, implying that the UK-based companies and the Swedish-based companies apply a similar tone in their ESG disclosure. This does not seem to be an unreasonable finding – given that ESG disclosure is a new concept and that much of the Swedish-based and the UK-based legislation, and interpretation of that legislation, is built on EU code. Further, the concept of ESG has been born out of the modern world – where many companies use the same type of communicative models regardless of their domicile. The lack of difference, even though the original language in one of the corpora studied is not English (assuming that the Swedish-based companies first write their ESG report in Swedish, and then have it translated), is also an interesting find. This suggests today's financial world is interconnected and conformist, which differs from Bondi and Yu's (2019) findings that companies with different cultural backgrounds disclose their ESG narratives differently. Ultimately, it might be, as postulated by Bahita (2008), that additional understanding of the texts would be needed to discern if there are any differences. Applying a text-mining approach alone may not be sufficient to reveal differences between the two corpora, and a qualitative method should also be used to fully understand the narratives.

6.2. Further research

Interesting areas of further research could be a comparison between the level of Optimism in corpora based in predominantly anglophone countries such as the US, Canada (English-speaking part of), Australia and the UK to determine if these differ or are aligned. Similarly, a split on industry could yield interesting results, for example, investigating if the level of Optimism in the language used for ESG disclosures is different for an automotive manufacturer, as compared to an agricultural company.

References

Albitar, K., Abdoush, T., & Hussainey, K. (2022). Do corporate governance mechanisms and ESG disclosure drive CSR narrative tones? *International Journal of Finance & Economics, 1– 15*. https://doi.org/10.1002/ijfe.2625

Bell, E., Bryman, A., & Harley, B. (2019). *Business research methods* (6th ed.). Oxford, UK: Oxford University Press.

Bahita, V. K. (2008). Genre analysis, ESP and professional practice. *English for Specific Purposes, 27 (2)*, 161-174. https://doi.org/10.1016/j.esp.2007.07.005

Bondi, M., & Yu, D. (2019). Textual Voices in Corporate Reporting: A Cross-Cultural Analysis of Chinese, Italian, and American CSR Reports. *International Journal of Business Communication, 56 (2)*, 173–197. https://doi.org/10.1177/2329488418784690

Dainow, J. (1966). The Civil Law and the Common Law: Some Points of Comparison. *The American Journal of Comparative Law, 15(3)*, 419–435. https://doi.org/10.2307/838275

Doshi, A.R., Dowell, G.W.S. and Toffel, M.W. (2013), How firms respond to mandatory information disclosure. *Strat. Mgmt. J., 34,* 1209-1231. https://doi.org/10.1002/smj.2055

Finegan, E. (2008). Language: its use and structure. *Thomson Wadsworth*.

Hellman, A., Patel, C., & Tsunogaya, N. (2021). Foreign-language effect and professionals' judgments on fair value measurement: Evidence from Germany and the United Kingdom. *Journal of Behavioural and Experimental Finance, 30,* 100478. https://doi.org/10.1016/j.jbef.2021.100478

Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P. (2022). Text Pre-processing for Text Mining in Organizational Research: Review and Recommendations. *Organizational Research Methods, 25(1)*, 114–146. https://doi.org/10.1177/1094428120971683

Hotho, A., Nürnberger, A., & Paass, G. (2005). A Brief Survey of Text Mining. LDV Forum, 20, 19-62.

Jones, M. J., Melis, A., Gaia, S., & Aresu, S. (2020). Impression Management and Retrospective Sense-Making in Corporate Annual Reports: Banks' Graphical Reporting

During the Global Financial Crisis. *International Journal of Business Communication, 57(4)*, 474–496. https://doi.org/10.1177/2329488417712010

Laskin, A. V., & Mikhailovna Nesova, N. (2022). The Language of Optimism in Corporate Sustainability Reports: A Computerized Content Analysis. *Business and Professional Communication Quarterly, 85(1)*, 80–98. https://doi.org/10.1177/23294906211065507

Lindquist, H., & Levin, M. (2018). Corpus Linguistics and the Description of English (2nd ed.). *Edinburgh University Press*.

McCarthy, M., & O'Keeffe, A. (2010). Historical perspective: What are corpora and how have they evolved? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 3–13). London, UK: Routledge.

Roy, A., & Mukherjee, P. (2022). Does National Culture Influence Corporate ESG Disclosures? Evidence from Cross-Country Study. *Vision, 0(0)*. https://doi.org/10.1177/09722629221074914

Saeed, J. I. (2016). Semantics (4th ed.). *Wiley Blackwell*.

Smith, V., Florence, K., & Maria, F. (2018). Semantics in cultural perspective overview. *Linguistics and Culture Review, 2(1)*, 24-31. https://doi.org/10.37028/lingcure.v2n1.9

Vaughan, E. and O'Keeffe, A. (2015). Corpus Analysis. *In The International Encyclopaedia of Language and Social Interaction* (eds K. Tracy, T. Sandel and C. Ilie). https://doi.org/10.1002/9781118611463.wbielsi168

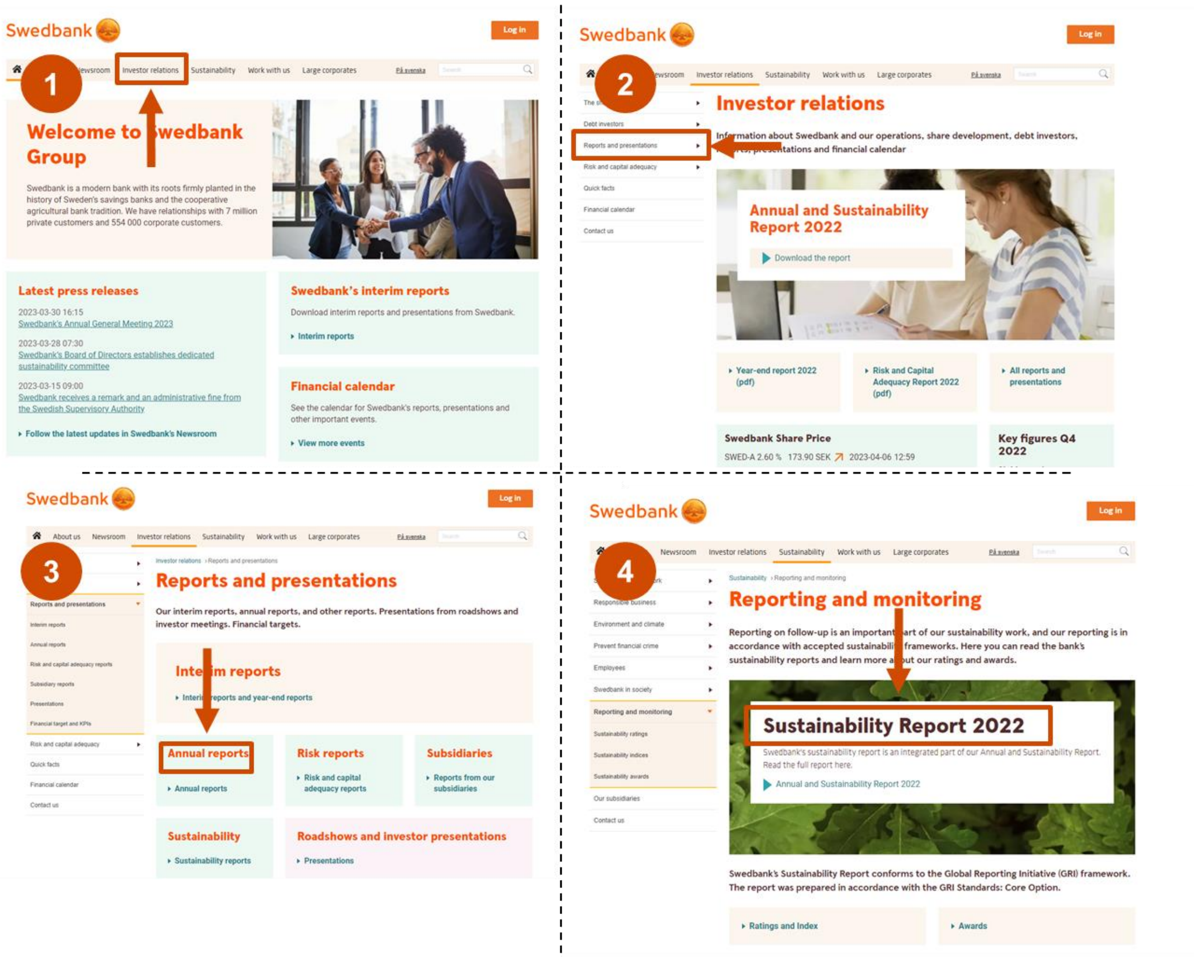
Yakut, I. (2022). The basics of corpus linguistics: An introduction for beginners. In H. Yilmaz, E. Yucel, & M. S. Ozturk (Eds.), A general perspective on foreign language teaching VI (pp. 176-190). Çizgi Kitabevi. Retrieved from https://www.cizgikitabevi.com/kitap/1564-yabanci-dil-ogretimine-genel-bir-bakis-vi

Appendix I

Downloading the ESG disclosures

The ESG disclosures were collected from the annual reports found on the "Investor" section of the listed companies' websites. The annual reports are located under the Financial Reporting section. The ESG disclosure is typically labelled as "ESG Disclosure", "ESG Report", or "Sustainability Report". Below outlines the process of extracting the ESG disclosure of the Swedish-based bank Swedbank (screen captures are as of the English version of their web page and are extracted as of 8th of April 2023, red arrows and red brackets added by author to highlight relevant fields).

1. **Accessing the investor relations**: All publicly traded companies are obliged to have an ‘Investor relations’ section on their homepage where they disclose data to the market.
2. **Reports and presentations:** The different reports and presentations made available to the public are summarized under a separate section.
3. **Annual reports:** The annual report section outlines the reports the company need to disclose once a year.
4. **Sustainability reporting:** Outlines the available sustainability reports.



Swedbank
Log in
Investor relations
Sustainability
Work with us
Large corporates
1
Welcome to Swedbank Group
Swedbank is a modern bank with its roots firmly planted in the history of Sweden's savings banks and the cooperative agricultural bank tradition. We have relationships with 7 million private customers and 554 000 corporate customers.
Latest press releases
2023-03-30 16:15
Swedbank's Annual General Meeting 2023
2023-03-28 07:30
Swedbank's Board of Directors establishes dedicated sustainability committee
2023-03-15 09:00
Swedbank receives a remark and an administrative fine from the Swedish Supervisory Authority
Follow the latest updates in Swedbank's Newsroom
Swedbank's interim reports
Download interim reports and presentations from Swedbank.
Interim reports
Financial calendar
See the calendar for Swedbank's reports, presentations and other important events.
View more events
2
Investor relations
Reports and presentations
Annual and Sustainability Report 2022
Download the report
Year-end report 2022 (pdf)
Risk and Capital Adequacy Report 2022 (pdf)
All reports and presentations
Swedbank Share Price
SWED-A 2.60 % 173.90 SEK 2023-04-06 12:59
Key figures Q4 2022
3
Reports and presentations
Our interim reports, annual reports, and other reports. Presentations from roadshows and investor meetings. Financial targets.
Interim reports
Interim reports and year-end reports
Annual reports
Risk reports
Risk and capital adequacy reports
Subsidiaries
Reports from our subsidiaries
Sustainability
Sustainability reports
Roadshows and investor presentations
Presentations
4
Reporting and monitoring
Reporting on follow-up is an important part of our sustainability work, and our reporting is in accordance with accepted sustainability frameworks. Here you can read the bank's sustainability reports and learn more about our ratings and awards.
Sustainability Report 2022
Annual and Sustainability Report 2022
Swedbank's Sustainability Report conforms to the Global Reporting Initiative (GRI) framework. The report was prepared in accordance with the GRI Standards: Core Option.
Ratings and Index
Awards

Appendix II

SPSS output per sub-dictionary

Praise

Group Statistics

	Domicile	N	Mean	Std. Deviation	Std. Error Mean
Praise_POS	SE	21	2,0981	2,08280	,45450
	UK	21	1,9852	2,00665	,43789

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
Praise_POS	Equal variances assumed	,018	,894	,179	40	,429	,859	,11286	,63112	-1,16269	1,38841
	Equal variances not assumed			,179	39,945	,429	,859	,11286	,63112	-1,16275	1,38846

Satisfaction

Group Statistics

	Domicile	N	Mean	Std. Deviation	Std. Error Mean
Satisfaction_POS	SE	21	,9424	1,75722	,38346
	UK	21	1,0519	1,18903	,25947

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
Satisfaction_POS	Equal variances assumed	,034	,856	-,237	40	,407	,814	-,10952	,46299	-1,04527	,82622
	Equal variances not assumed			-,237	35,140	,407	,814	-,10952	,46299	-1,04931	,83027

Inspiration

Group Statistics

	Domicile	N	Mean	Std. Deviation	Std. Error Mean
Inspiration_POS	SE	21	4,0357	1,83625	,40070
	UK	21	3,1995	1,63873	,35760

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
Inspiration_POS	Equal variances assumed	,257	,615	1,557	40	,064	,127	,83619	,53707	-,24926	1,92164
	Equal variances not assumed			1,557	39,493	,064	,127	,83619	,53707	-,24969	1,92208

Blame

Group Statistics

	Domicile	N	Mean	Std. Deviation	Std. Error Mean
Blame_NEG	SE	21	,3514	,32197	,07026
	UK	21	,4229	,94485	,20618

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance		Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p			Lower	Upper
Blame_NEG	Equal variances assumed	,955	,334	-,328	40	,372	,745	-,07143	,21783	-,51167	,36881
	Equal variances not assumed			-,328	24,583	,373	,746	-,07143	,21783	-,52043	,37758

Hardship

Group Statistics

	Domicile	N	Mean	Std. Deviation	Std. Error Mean
Hardship_NEG	SE	21	3,6462	2,49344	,54411
	UK	21	4,7229	2,95121	,64401

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance		Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p			Lower	Upper
Hardship_NEG	Equal variances assumed	,143	,707	-1,277	40	,104	,209	-1,07667	,84309	-2,78062	,62729
	Equal variances not assumed			-1,277	38,915	,105	,209	-1,07667	,84309	-2,78210	,62877

Denial

Group Statistics

	Domicile	N	Mean	Std. Deviation	Std. Error Mean
Denial_NEG	SE	21	3,9257	9,99259	2,18056
	UK	21	3,0414	2,79074	,60899

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance		Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p			Lower	Upper
Denial_NEG	Equal variances assumed	1,257	,269	,391	40	,349	,698	,88429	2,26400	-3,69144	5,46001
	Equal variances not assumed			,391	23,101	,350	,700	,88429	2,26400	-3,79803	5,56660

Appendix III

Development per year of sub-dictionary means

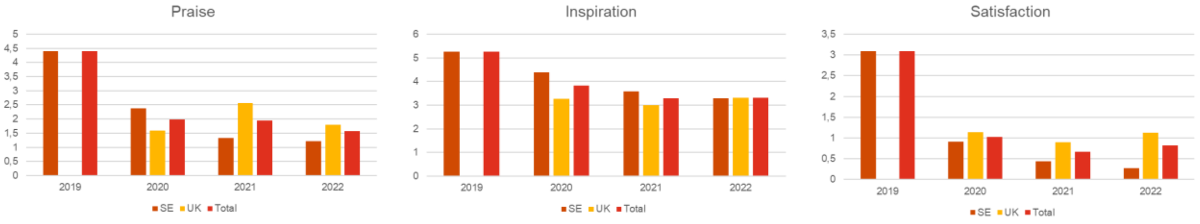


Figure 4. Visualisation of the three positively correlated sub-dictionaries, Praise, Inspiration and Satisfaction and their development over the period

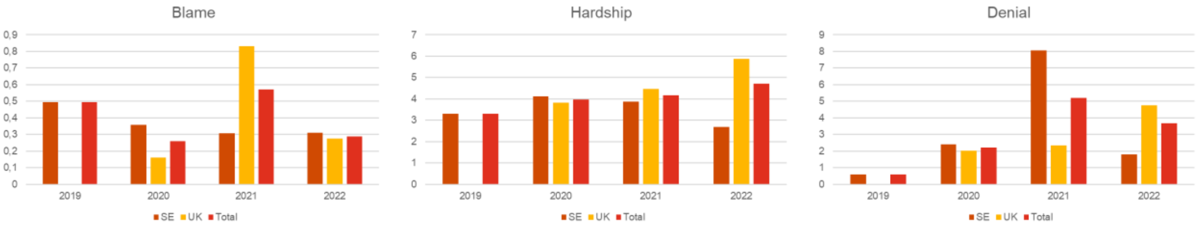


Figure 5. Visualisation of the three negatively correlated sub-dictionaries, Blame, Hardship and Denial and their development over the period b