



A thematically oriented analysis of the Financial Services Annual Reports (FinSerAR) Corpus: the UK financial services' narrative towards Brexit

1/11/2018

Vasiliki Simaki

ESRC Centre for Corpus Approaches to Social Science (CASS), Linguistics and English Language Department

Summary

- Introduction to financial text & analysis
- The FinSerAR corpus & the Brexit-related content
- Brexit-related content analysis
 - Analysis of the Brexit-related terms context
 - Analysis of the Brexit-related content
- Stance identification in the Brexit-related content
- Conclusions & future work

Financial text

- Financial reporting documents of different types
 - Text data from financial reporting narratives (annual reports, earnings statements)
 - Text data from financial articles from the press or other sources (websites, social media)
 - Text data from press releases
 - Spoken data from press conferences (earning calls & presentations)
 - Spoken data from other financial communication
- Differences in structure, content & context → same purpose: the publication of the company's financial status to an expert audience & to the society

Topics of interest

- Genre analysis: diachronic evolution of the text type (length, sections, content, style), identify motives, patterns, rules (Beattie et al. 2008, Bhatia 2008)
- Readability studies based on large text collections (Li 2008), analysis & comparison of bilingual corpora (Lang & Stice-Lawrence 2015)
- Stylistic analysis: writing style of specific sections, lexical choices, word frequencies → profile of the narrative (Rutherford 2005, Wang et al. 2012)
- Thematic studies, e.g. the 2008-2011 financial crisis in financial reporting (Dragsted 2014)
- The use of ICT tools in financial reporting (Crawford Camiciottoli 2013)
- Comparative studies using general reference corpora such as BNC & LOB (Tian & Liang 2011)
- Study specific linguistic phenomena in financial reporting, e.g. metaphors (Charteris-Black, J. & Musolff, A. 2003, Charteris-Black, J., & Ennis, T. 2001)
- Identify good/bad reporting practices (Goel et al. 2010)

Annual Reports (ARs)

- Comprehensive report on a company's activities throughout the preceding year
- Sub-genre of organisational communication
- Complex multimodal document
 - Consisting of different narratives and sections (US & EU ARs)
 - Aiming to different audiences at the same time
 - Dual informational-promotional function

(Bhatia 2004, 2008)

UK financial services' narrative towards Brexit

“70% of the estimated costs of the UK’s EU exit will be incurred by five sectors in the UK: **financial services**, automotive, agriculture, food and drink, consumer goods, and chemicals and plastics”*

(Oliver Wyman and Clifford Change 2018,

<http://www.oliverwyman.com/our-expertise/insights/2018/mar/red-tape-cost-brexit.html>)

- Focus on financial services’ annual reports
- Corpus-based methodology → analyse how these companies refer to Brexit in their recent annual reports

* estimated costs from trade barriers to UK and EU27 firms

Corpus description

- Four independent British financial companies: Barclays PLC, HSBC Holdings PLC, Lloyds Banking Group, Royal Bank of Scotland
- Foreign bank incorporated in the UK: Santander UK PLC
- 2015, 2016 & 2017 ARs
- Download the ARs in PDF & convert to txt using the CFIE-FRSE Web Tool (<https://cfie.lancaster.ac.uk:8443/>) implemented within the Corporate Financial Information Environment (CFIE) project @Lancaster University (<http://ucrel.lancs.ac.uk/cfie/>)
- The Financial Services Annual reports (FinSerAR) corpus

The FinSerAR Corpus

- About 3.2 million words

ARs	Barclays	HSBC	Lloyds	RBS	Santander UK	Total
2015	85,842	288,150	215,678	245,769	221,877	1,057,316
2016	247,058	185,697	205,343	266,955	171,629	1,076,682
2017	221,197	185,739	198,406	251,592	162,941	1,019,875
Total:	554,097	659,586	617,427	764,316	556,447	3,151,873

The Brexit-related content

- Manually extract all the content related to the 2016 UK referendum
- Brexit-related terms: *Brexit, referendum, exit, leav* the EU**
- Term & surrounding context
- One sentence to one or more paragraphs
- Manual selection → small but accurate data set
- 28,645 words in total
 - 2015 subset: 2,270 words
 - 2016 subset: 15,239 words
 - 2017 subset: 11,136 words
- Each text chunk → annotated with the AR's section

The size of the Brexit subset

Financial company	2015	2016	2017	Total:
Barclays PLC	258	5,510	2,982	8,750
HSBC Holdings PLC	216	1,406	1,102	2,724
Lloyds Banking Group	645	2,406	2,022	5,073
Royal Bank of Scotland	835	4,658	4,360	9,844
Santander UK PLC	316	1,259	679	2,254
Total:	2,270	15,239	11,136	28,645

Analysis of the Brexit-related terms context

Keyword	2015 ARs	2016 ARs	2017 ARs
Brexit	1	1	58
exit	15	44	50
referendum	28	127	23
leav* the EU*	4	49	18

- Observe the context around the Brexit-related terms, AntConc (Anthony 2014)
- Identify differences between the three yearly subsets
- Discuss findings

Brexit term examples

- *We have assessed the potential consequences for our business of the UK leaving the EU (**Brexit**), as well as the potential impact of market instability in the lead up to the referendum and in any implementation period following a potential ‘leave’ vote. (2015)*
- *In particular, we highlight the threat of populism impacting policy choices in upcoming European elections, possible protectionist measures from the new US administration impacting global trade, uncertainties facing the UK and the EU as they enter **Brexit** negotiations, and the impact of a stronger dollar on emerging economies with high debt levels. (2016)*
- *We continued with our **Brexit** preparations to ensure that Barclays can preserve our access to the EU markets for our customers and clients. (2017)*
- *Additional structural changes to the Group’s operations will also be required as a result of **Brexit**. (2017)*

2017 subset
result
developments
Group
uncertainty
transformation
required
including
restructuring
political
impact

Exit term collocates & examples

- *In the UK, the referendum on EU membership gives rise to some political uncertainty and raises the possibility of a disruptive and uncertain **exit** from the EU, with attendant consequences for investment and confidence. (2015)*
- *Similarly, the impact of the planned **exit** of the UK from the EU could potentially have an impact on our ability to hire and retain key employees. (2016)*
- *We are making comprehensive plans for the UK's planned **exit** from the EU and we believe we will provide an uninterrupted service to our clients, consumers and other stakeholders during and after the transition. (2017)*
- *An uncertain UK and global economic outlook and uncertainty relating to EU **exit** negotiations have the ability to impact the Commercial Banking portfolios. (2017)*

2015 ARs	2016 ARs	2017 ARs
UK	EU	EU
EU	UK	UK
Union	<i>risks</i>	negotiations
European	<i>referendum</i>	<i>impacts</i>
<i>uncertainty</i>	outcome	<i>risks</i>
supporting	<i>possible</i>	<i>potential</i>
relating	<i>impact</i>	<i>impact</i>
<i>referendum</i>	Group	countries
<i>potential</i>	scenarios	<i>uncertainty</i>
occurs	<i>risk</i>	triggered

Referendum term collocates & examples

- *We continue to deal with a range of uncertainties in the external environment, including those caused by the **referendum** on the UK's continuing membership of the European Union. (2015)*
- *As a result of the **referendum** outcome and to manage the impact of uncertainty caused by the **referendum** process and ensuing economic concerns, detailed EU exit portfolio assessments were undertaken to understand potential impacts on the Bank's credit risk profile and to assess the potential need for any changes to Group risk appetite. (2016)*
- *This uncertainty is compounded by the UK's decision to leave the EU following the outcome of the EU **Referendum** which may result in further changes to the prudential and regulatory framework applicable to the Group. (2016)*
- *The Group is subject to political risks, including economic, regulatory and political uncertainty arising from the **referendum** on the UK's membership of the European Union which could adversely impact the Group's business, results of operations, financial condition and prospects. (2017)*
- *Political risks continue to evolve with the UK's vote to leave in the EU **referendum** creating significant economic, political and regulatory uncertainty. (2017)*

2015 ARs	2016 ARs	2017 ARs
UK	EU	EU
EU	following	UK
membership	membership	outcome
risks	result	political
outcome	outcome	June
lead	June	following
hold	impact	result
rise	post	uncertainty
potential	after	Scottish
period	risk	monitored

leav the EU** term collocates & examples

- *Following the referendum in June 2016, in the event that there is a vote in favour of **leaving the EU**, a period of negotiation is likely, widely anticipated to be around two years, with unpredictable implications on market conditions. (2015)*
- *In light of these potential developments as well as the impact of the UK's decision to **leave the EU** following the result of the EU Referendum, there remains uncertainty as to the rules which may apply to the Group going forward. (2016)*
- *Finally, there were the challenges presented by emerging economic and political risks, notably those associated with the EU Referendum and the subsequent vote by the UK to **leave the EU**. (2016)*
- *This had been prompted by the rising level of personal debt in the UK and concerns of weaker growth and higher inflation resulting from the country's vote to **leave the European Union**. (2017)*
- *In addition, it is possible (although of low likelihood) that a disorderly termination of the Article 50 process could occur, resulting in the UK **leaving the EU** before 29 March 2019. (2017)*

2016 ARs	2017 ARs
EU	EU
UK	UK
vote	decision
decision	resulting
following	March
result	following
outcome	before
June	uncertainty
favour	terms
expected	significant

Analysis of the Brexit-related content

- Statistical analysis → observe & highlight differences between the three subsets
- Extract the word lists in terms of descending word frequency of each subset
- List lexical items that appear more than 0.25% of all word tokens in each subset → 94 words (stop words incl.)
- Chi-square (χ^2) test: determine the population distribution of each subset in the corpus
 - p-value <0.05 → 32 statistically significant items
- Post-hoc test: Bonferroni correction → set a lower significance level, <99% of confidence
 - p-value =0.0001 & critical value = ± 3.47 → divide the familywise error rate (0.05) by the number of tests (96)

Words	2015	2016	2017	p-value
ability	0.31%	0.10%	0.15%	0.048
any	0.57%	0.24%	0.20%	0.005
as	1.01%	0.78%	1.21%	0.002
Barclays	0.04%	0.41%	0.27%	0.006
Brexit	0.04%	0.01%	0.52%	0.000
could	0.62%	0.21%	0.34%	0.002
developments	0.44%	0.18%	0.22%	0.049
effect	0.22%	0.07%	0.04%	0.019
European	0.62%	0.28%	0.40%	0.019
exit	0.66%	0.26%	0.40%	0.006
following	0.13%	0.31%	0.15%	0.021
for	0.35%	0.85%	0.86%	0.040
geopolitical	0.26%	0.04%	0.12%	0.001
government	0.31%	0.06%	0.09%	0.001
Group	0.88%	0.81%	1.12%	0.036
leave	0.09%	0.28%	0.13%	0.015
membership	0.75%	0.16%	0.05%	0.000
on	1.72%	1.28%	0.91%	0.001
or	1.41%	0.56%	0.59%	0.000
ratings	0.35%	0.10%	0.08%	0.002
RBS	0.48%	0.12%	0.13%	0.000
referendum	1.28%	0.85%	0.22%	0.000
regulatory	0.44%	0.28%	0.51%	0.011
risk	0.70%	0.80%	0.54%	0.041
risks	0.70%	0.36%	0.47%	0.049
structural	0.04%	0.14%	0.24%	0.040
UK	2.16%	1.38%	1.61%	0.015
union	0.48%	0.12%	0.19%	0.001
vote	0.26%	0.20%	0.03%	0.000
was	0.00%	0.28%	0.09%	0.000
with	0.13%	0.51%	0.73%	0.001
would	0.26%	0.09%	0.07%	0.021

Words	2015	2016	2017
ability	2.244	-1.674	0.471
any	3.186	-0.308	-1.450
as	0.234	-3.432	3.383
Barclays	-2.467	2.690	-1.387
Brexit	-1.796	-8.008	9.193
could	2.971	-2.778	1.198
developments	2.394	-1.270	-0.027
effect	2.696	-0.075	-1.417
European	2.213	-2.344	1.173
exit	2.667	-2.556	1.139
following	-1.046	2.784	-2.270
for	-2.548	0.665	0.731
geopolitical	2.977	-2.927	1.347
government	3.588	-1.900	-0.043
Group	-0.299	-2.346	2.567
leave	-1.318	2.870	-2.207
membership	7.175	-0.294	-3.675
on	2.534	1.848	-3.296
or	4.804	-1.836	-0.782
ratings	3.578	-0.717	-1.248
RBS	4.106	-1.477	-0.763
referendum	3.980	4.703	-7.020
regulatory	0.454	-2.971	2.790
risk	0.082	2.382	-2.483
risks	2.092	-1.890	0.775
structural	-1.526	-1.452	2.332
UK	2.530	-2.173	0.822
union	3.610	-2.284	0.337
vote	1.726	2.971	-3.998
was	-2.117	3.988	-2.909
with	-2.870	-1.292	2.913
would	2.752	-0.526	-0.986

Findings

- 2015 different words: *government , or, ratings, RBS, referendum* , Union*
- 2016: *referendum* , was*
- 2017: *Brexit*
- 2015 reports VS 2016 & 2017 reports
- Low number of the different words → not many differences between the 2015, 2016 & 2017 reports

The 20 highest ranked keywords of the 2015, 2016, 2017 reports, and overall

- BNC as a reference corpus

2015 ARs	2016 ARs	2017 ARs	all ARs
EU	EU	EU	EU
UK	UK	UK	UK
referendum	referendum	Brexit	referendum
RBS	risk	Group	impact
exit	impact	regulatory	risk
risks	Barclays	risks	Group
membership	Group	exit	regulatory
impact	risks	impact	risks
geopolitical	volatility	risk	Brexit
regulatory	regulatory	uncertainty	Barclays
risk	exit	Barclays	exit
uncertainty	potential	including	uncertainty
including	including	referendum	including
Group	uncertainty	financial	volatility
ratings	credit	geopolitical	RBS
developments	economic	structural	potential
European	markets	volatility	credit
changes	Eurozone	RBS	financial
credit	RBS	European	economic
potential	financial	credit	global

The overall keywords categorised in terms of their use in the text

Brexit-related terms	Charged words	Neutral words
EU	risk/risks	Group
UK	impact	Barclays
referendum	regulatory	including
Brexit	uncertainty	RBS
exit	volatility	potential
		credit
		financial
		economic
		global

- *risk/s, impact & uncertainty*: describe in a negative way the companies' reaction & impression about the 2016 UK referendum
- *volatility*: 2016 & 2017, emphasising the instability that the referendum's outcome has created to the companies, the clients & the market
- *regulatory*: in that context they refer to the transformations, reforms & changes that need to be done throughout the Brexit process

Stance identification in the Brexit-related content

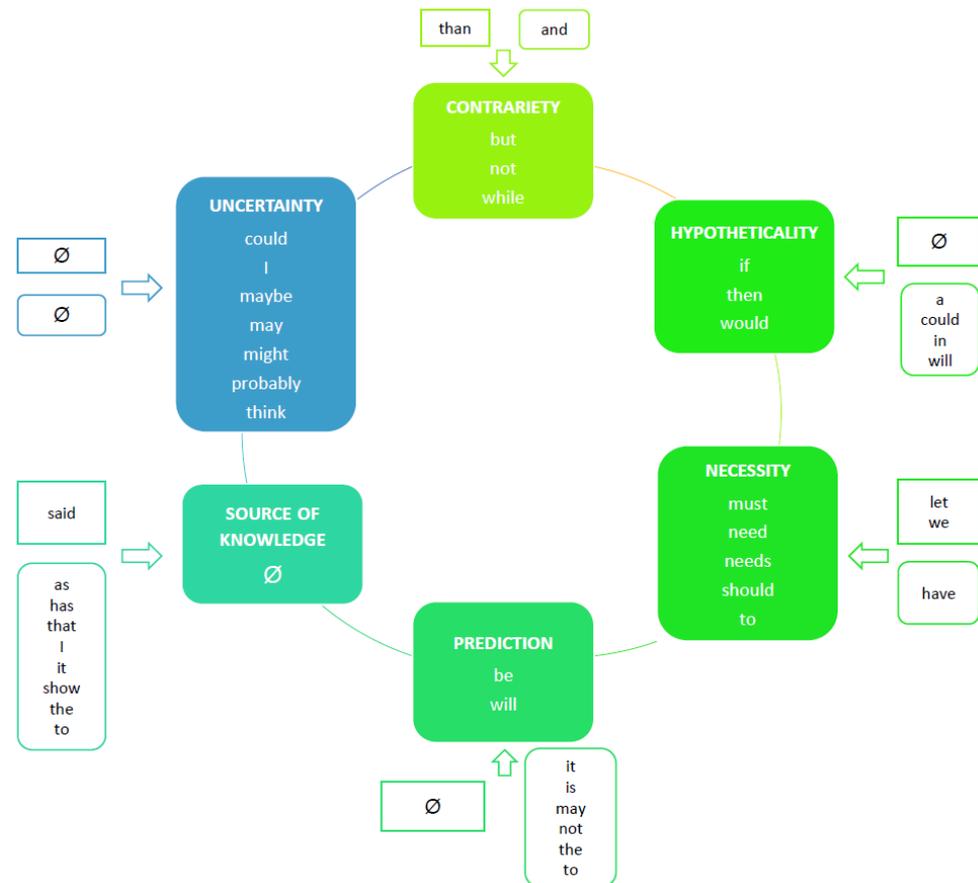
Subjective position of the companies towards Brexit → stance-taking

- Simaki et al. (2017) stance framework, ten notional categories
- Corpus of opinionated texts, the Brexit Blog Corpus (BBC)
- Analytical protocol & interface (ALVA, Kucher et al. 2017) for the annotations
- Utterances independently annotated by two annotators
- Evaluate this framework in a different text type but same thematic



The stance markers in the BBC (Simaki et al. 2019)

- Statistically significant items for six stance categories
- Most frequent items from the meta-annotation process
- Stance-related items that do not unambiguously evoke a specific stance but are parts of larger chunks of wordings in stance constructions
- Well-established stance-related markers were confirmed as significant stance constructions in the BBC



Stance analysis of the Brexit-related content

- Statistical analysis of the FinSerAR corpus → stance markers occurrence in the Brexit- & non Brexit-related content
- Extract the stance markers in both sets
- Chi-square (χ^2) test, p-value <0.05 → 20 statistically significant items
- Post-hoc test: Bonferroni correction, p-value = 0.0001 & critical value = ± 3.83 →
divide the familywise error rate (0.05) by the number of tests (40)

Stance category	Stance word	Brexit-related content	non Brexit-related content	Total	p-value	
Contrariety	but	50	2,191	2,241	0.000	
	not	45	7,335	7,380	0.002	
	while	16	807	823	0.004	
	than	13	3,605	3,618	0.000	
	and	1,343	11,1627	11,2970	0.000	
Hypotheticality	if	14	2,401	2,415	0.056	
	then	3	493	496	0.418	
	would	27	1,645	1,672	0.006	
	and	377	39,691	40,068	0.711	
	could	84	1,684	1,768	0.000	
	in	773	64,714	65,487	0.000	
	will	74	5,533	5,607	0.006	
Necessity	must	0	549	549	0.021	
	need	16	511	527	0.000	
	needs	4	663	667	0.341	
	should	10	756	766	0.325	
	to	949	78,653	79,602	0.000	
	let	2	175	177	0.815	
	we	153	13,299	13,452	0.034	
	have	160	7,670	7,830	0.000	
Prediction	be	118	11,021	11,139	0.277	
	will	74	5,533	5,607	0.006	
	it	47	5,255	5,302	0.588	
	is	138	22,983	23,121	0.000	
	may	117	4,876	4,993	0.000	
	not	45	7,344	7,389	0.002	
	the	2,554	174,412	176,966	0.000	
	to	949	78,653	79,602	0.000	
	Source of knowledge	said	0	26	26	0.616
		as	277	20,533	20,810	0.000
has		88	6,250	6,338	0.000	
that		123	14,727	14,850	0.102	
I		16	2,308	2,324	0.181	
it		47	5,255	5,302	0.588	
show		1	217	218	0.449	
the		2,554	174,412	176,966	0.000	
to		949	78,653	79,602	0.000	
Uncertainty		could	84	1,684	1,768	0.000
	I	16	2,308	2,324	0.181	
	maybe	0	0	0	-	
	may	117	4,876	4,993	0.000	
	might	5	330	335	0.316	
	probably	0	5	5	0.826	
	think	2	43	45	0.016	
	Total:		28,645	2,958,492	2,987,137	

Stance category	Stance word	Brexit content	non Brexit content
Contrariety	but	6.182092	-6.18209
	not	-3.08192	3.081919
	while	2.900434	-2.90043
	than	-3.7032	3.703196
	and	8.082092	-8.08209
Hypotheticality	could	16.36646	-16.3665
	would	2.752736	-2.75274
	will	2.775084	-2.77508
	in	5.879583	-5.87958
	must	-2.30577	2.305766
	we	2.128348	-2.12835
	Necessity	need	4.89326
to		6.844133	-6.84413
have		9.859775	-9.85977
Prediction	is	-5.67148	5.671483
	may	10.04572	-10.0457
	will	2.775084	-2.77508
	not	-3.09036	3.090362
	the	21.55209	-21.5521
Source of knowledge	to	6.844133	-6.84413
	as	5.527933	-5.52793
	has	3.512382	-3.51238
	the	21.55209	-21.5521
	to	6.844133	-6.84413
Uncertainty	could	16.36646	-16.3665
	may	10.04572	-10.0457
	think	2.39922	-2.39922

Results & future steps

- Contrariety: *but*
- Necessity: *need*
- Hypotheticality/Uncertainty: *could*
- Prediction/Uncertainty: *may*
- *the, to, and, have, as, in* : parts of bigger stance constructions

- Brexit-related content → less neutral in terms of stance in comparison to the rest of the FinSerAR corpus
- Objective/neutral language in the reports VS charged & stanced discourse about Brexit
- Financial services' interests & strong opinions towards politico-economic developments
- Influence clients, market, society & create an atmosphere of uncertainty, generic negative narrative

- What about sentiment in the corpus?
- Expand Brexit-related content with data from other companies' ARs

References

-
- Anthony, L. 2014. AntConc (Version 3.4.3) [Computer Software], Waseda University, Tokyo, Japan. Available at <http://www.laurenceanthony.net/>
- Beattie, V., Dhanani, A., & Jones, M. J. (2008). Investigating presentational change in UK annual reports: A longitudinal perspective. *The Journal of Business Communication* (1973), 45(2), 181-222.
- Bhatia, V. K. (2008). Genre analysis, ESP and professional practice. *English for specific purposes*, 27(2), 161-174.
- Bhatia, V. (2004). *Worlds of written discourse: A genre-based view*. A&C Black.
- Charteris-Black, J., & Musolff, A. (2003). 'Battered hero' or 'innocent victim'? A comparative study of metaphors for euro trading in British and German financial reporting. *English for Specific Purposes*, 22(2), 153-176.
- Charteris-Black, J., & Ennis, T. (2001). A comparative study of metaphor in Spanish and English financial reporting. *English for specific purposes*, 20(3), 249-266.
- Crawford Camiciottoli, B. (2013). *Rhetoric in financial discourse: A linguistic analysis of ICT-mediated disclosure genres* (pp. 1-212). Editions Rodopi BV.
- Dragsted, B. (2014). A case study of letters to shareholders in annual reports before, during and after the financial crisis. *LSP Journal-Language for special purposes, professional communication, knowledge management and cognition*, 5(2).
- Goel, S., Gangolly, J., Faerman, S. R., & Uzuner, O. (2010). Can linguistic predictors detect fraudulent financial filings?. *Journal of Emerging Technologies in Accounting*, 7(1), 25-46.
- Kucher, K., Paradis, C., Sahlgren, M., & Kerren, A. (2017). Active Learning and Visual Analytics for Stance Classification with ALVA. *ACM Transactions on Interactive Intelligent Systems*. ACM Press.

-
- Lang, M., & Stice-Lawrence, L. (2015). Textual analysis and international financial reporting: Large sample evidence. *Journal of Accounting and Economics*, 60(2-3), 110-135.
- Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, 45(2-3), 221-247.
- Rutherford, B. A. (2005). Genre analysis of corporate annual report narratives: A corpus linguistics–based approach. *The Journal of Business Communication (1973)*, 42(4), 349-378.
- Simaki, V., Paradis, C., Skeppstedt, M., Sahlgren, M., Kucher K., & Kerren A. 2017. Annotating speaker stance in discourse: the Brexit Blog Corpus, *Corpus Linguistics and Linguistic Theory*. DOI: 10.1515/cllt-2016-0060
- Simaki, V., Paradis, C., Kerren, A. (2019). A two-step procedure to identify lexical elements of stance constructions in discourse from political blogs. In *Corpora* (accepted).
- Tian, X. and Liang, H. (2011). A corpus based study on stylistic analysis of auditing reports. *Market Modernization* 642, 136-138.
- Wang, H., Li, L., & Cao, J. (2012). Lexical features in corporate annual reports: a corpus-based study. *European Journal of Business and Social Sciences*, 1(9), 55-71.

