

PASCUAL PÉREZ-PAREDES

A Keyword Analysis of the 2015 UK Higher Education Green Paper and the Twitter Debate

1. Introduction

The UK ‘Higher education (HE): teaching excellence, social mobility and student choice’ Green Paper (henceforth HEGP) was released in November 2015. One of the flagships of this initiative is the implementation of the teaching excellence framework (TEF), which will allegedly give power to the UK government to monitor and assess the quality of teaching in English universities. According to *Times Higher Education*,¹ the new TEF seeks to ensure that students receive an “excellent” teaching experience that encourages “original thinking, drives up engagement and prepares them for the world of work, build a culture where teaching has equal status with research, provide students with the information they need to judge teaching quality and recognize institutions that do the most to welcome students and include a clear set of outcome-focused criteria and metrics”. HEGP consultation ended on 15 January 2016².

The UK Parliament³ defines green papers as consultation documents produced by the government whose aim is to allow people inside and outside parliament to give feedback on policy or legislative proposals. The *Oxford English Dictionary* defines Green Papers as a preliminary report of government proposals published to stimulate

1 *Times Higher Education*. URL: <<https://www.timeshighereducation.com/news/teaching-excellence-framework-tef-everything-you-need-to-know>>

2 URL: <<https://www.gov.uk/government/consultations/higher-education-teaching-excellence-social-mobility-and-student-choice>>

3 URL: <<http://www.parliament.uk/site-information/glossary/green-papers/>>

discussion. In a way, green papers streamline the response and the reaction of different stakeholders to the questions and issues therein. Unsurprisingly, green papers give the Government a unique opportunity to manipulate those questions and issues to their advantage.

The use of social media is opening up new ways to voice criticism and express opinions in ways that were simply not possible a decade ago. The Academia is not oblivious to this trend. The number of scholars using social media, and, in particular, Twitter, is on the up. Although the implications for researchers and scholars are largely under-researched, an estimate in 2011⁴ claimed that 1 in 40 scholars used Twitter, and a more recent survey in Kashmir (Nazir *et al.* 2016) found that 30% of researchers in the social sciences in one single University used social networking tools, and 90% of them considered it important in their careers.

Using corpus-aided discourse analysis methods (Baker *et al.* 2008, 2013; Pérez-Paredes *et al.* 2017), together with social data mining techniques (Teh *et al.* 2015), this chapter sets out to examine how the analysis of keywords can help us understand how language use contributes to building the underlying discourses (Baker 2011) in the HEGP and the Twitter debate on the new TEF. Specifically, I will look at how keywords can reveal underlying discourse in both the Green Paper and the one found in social media at two different points in time: March and November 2016. Two research questions will be addressed in this paper:

1. What can Part of Speech (POS) keyword analysis reveal about power discourse in the HEGP?
2. How can keywords help us understand the Twitter debate on the TEF?

While Section 2 presents the POS keyword analysis of HEGP, Section 3 will discuss the method and the results of the keyword analysis of two Twitter datasets where the HEGP was debated. Section 4 offers a

4 URL:<<https://www.timeshighereducation.com/blog/weird-and-wonderful-world-academic-twitter>>

discussion of the results and Section 5 some conclusions, limitations and future research avenues.

2. POS keyword analysis of the Green Paper

2.1. Corpus linguistics, keyword analysis and discourse analysis

The use of keywords to investigate discourse has gained momentum over the last decade. Among other topics, researchers have examined the representation of Muslims in the UK press (Baker *et al.* 2013), the differences in the UK press during the Major and Blair governments (Jeffries/Walker 2012), the representation of immigrants in UK legal and administrative texts (Pérez-Paredes *et al.* 2017), or the representation of migrants, asylum seekers and refugees in the UK and Italian press (Taylor 2014). In this chapter, I will examine keywords as a way to investigate language use and gain further understanding of how the HEGP and the Twitter debates shape and reflect discourse (Stubbs 2007; Taylor in press) and how keyword analysis can be instrumental in gaining further understanding of how power and manipulation are constructed in the area of HE policy by different stakeholders. As I will explain later, two different approaches to keyword analyses will be used to research discourse in the context of this paper: POS keyword analysis and frequency-led keyword mining. I will get back to them in the following paragraphs.

O'Halloran (2010) and Taylor (in press) have suggested that two keyword conceptualization traditions have co-existed in the past. One is influenced by cultural studies traditions, and sees these words as the body of meanings of the practices that are central to our societies and institutions. The second tradition is embodied by corpus linguistics research methodology, one of its empirical principles being that "repeated events are significant" (Stubbs 2007: 130). In this light, the clustering of lexical items reveals different co-textual

environments that are built upon co-collocation and colligation (Pace-Sigge 2013). Frequency in a corpus or in a text is, in this way, observable evidence of probability in the system, therefore “unique events can be described only against the background of what is normal and expected” (Stubbs 2007: 130). Keywords, and, particularly, keyness (Scott/Tribble 2006), identify the lexical items that characterize a text or a whole corpus. In turn, Part of Speech (POS) keywords identify those POS tags that characterize a text or a POS-annotated corpus, which provides avenues for an interpretation of keywords that integrate both lexical as well as morphological layers of language.

The role of frequency, however, has been somewhat neglected in other research traditions. Fairclough’s (2001) work on power and language integrated the analysis of vocabulary and semantics at the lowest levels of study and interpretation of a text. In his framework, these two levels are predominantly ignored in favour of higher levels of analysis involving, among others, the role of situation of context in discourse interpretation. The assumption in this model is that words are used by speakers across the board non-contentiously as if word realization wasn’t conditioned and constrained by all previous textual uses and cotexts. Sinclair (2004: 120) criticized Fairclough (1999) on the grounds that one cannot take on board some of the collocations of a given term, that is, its behaviour in the texts produced by a community of users, and discard the ones which are not relevant to our interpretation of the text. Sinclair went on to say that the “citation of ‘used language’ proves nothing in itself about language unless the process of selectivity that is inevitable in such circumstances is controlled”. Corpus linguistics can provide such control, and corpus-driven evidence can help to show the “ideological trappings of a word or phrase”. This is precisely where keyword analysis may contribute to reveal the ideology of a given text or group of texts.

2.2. Part of speech keyword analysis of the HEGP

2.2.1. Method

By looking at the vocabulary used in a text or in a corpus through keyword analysis we accept the notion that repeated patterns of use are “widely shared in a discourse community” (Baker 2011: 13). POS keyword analysis reveals the central role played by some word categories in a text, providing an automatic profile of the part of speech (POS) tags that are significantly more frequent in text A when compared with reference corpus B. In the context of this research, we used the British National Corpus written sampler (BNCWS) as the contrast basis for our analysis. BNCWS is a collection of written material of 1,002,821 words, originally compiled to mirror the composition of the full written BNC (90 M)⁵. The HEGP consists of 33,099 words. The figures in brackets in this chapter represent the Log-likelihood (LL) value of the POS keyword analysis or alternatively the LogDice measure of the collocation strength between two lexical items. Items with a LL value over 6.63 are statistically significant as this is the cut-off for 99% confidence of significance. Wmatrix (Rayson 2008) was used for the POS keyword analysis and Sketch Engine (Kilgarriff *et al.* 2014) for the exploration of the collocational profiles of some of the lexical items. As for the tagset, Wmatrix is a web interface to the USAS and CLAWS corpus annotation tools and makes use of the UCREL CLAWS7 tagset⁶.

2.2.2. Results

Table 1 offers the most statistically significant POS tags in HEGP. Some of these (LL => 6.63) can be attributed to the repetitive nature of the terminology used in the HEGP. In fact, some of these features are so frequent that their presence is bordering 100% confidence interval. Comparative adjectives (687.99) such as *Higher* in *Higher Education*, proper nouns (210.55) such as the *DAPs* acronym in *Degree Awarding Powers*, textual numeric reference (171.04) such as *2* in *Chapter 2*, textual alphabetical intra-reference (83.91) such as *B* in *Part B*, dates and academic years (72.76) such as *2017/2018*,

5 For more information visit
<<http://ucrel.lancs.ac.uk/bnc2sampler/sampler.htm>>

6 URL: <<http://stig.lancs.ac.uk/wmatrix3/claws7tags.html>>

emerge as some of the most distinctive features in the HEGP when compared with BNCWS. I will not discuss these uses as they reflect some of the expected conventions of the genre of green papers (Fairclough 2003).

	POS tag	FreqHEG P	Freq BNCWS	LL		POS tag	FreqHEG P	Freq BNCWS	LL
1	JJR	402	2061	687.99	17	BCL21	32	158	56.56
2	NN2	2945	55665	469.45	18	FU	41	277	53.76
3	VVI	1378	24649	273.7	19	REX21	38	277	45.79
4	IF	609	8765	234.73	20	JK	36	265	42.9
5	NP2	81	231	210.55	21	VD0	86	1262	31.55
6	TO	826	14041	196.58	22	CST	307	6409	30.17
7	MC	761	13154	171.04	23	VBR	267	5435	30.06
8	VV0	661	11012	168.12	24	CC	1291	32450	27.29
9	VM	795	14301	154.59	25	FO	114	2050	22.2
10	NN1	5955	147395	151.91	26	CSW	30	325	20.78
11	PPIS2	263	3307	137.74	27	RRQ	106	2048	15.35
12	ZZ1	143	1698	83.91	28	II31	54	878	14.83
13	VVG	669	13564	76.88	29	RRT	13	111	12.89
14	MF	32	113	72.76	30	DD1	321	7677	11.76
15	DD2	153	2124	64.39	31	RRQV	9	69	10.22
16	VBI	378	7062	63.27	32	DAR	48	875	8.87

Table 1. POS keyword analysis.

Three groups of POS tags are prominently significant in our discussion of how power manifests itself in the HEGP: common nouns, the expression of finality, and modality. Below I offer a discussion of these uses together with, when convenient, collocational and colligational profiles of some of the most frequent realizations of these POS tags. These profiles expose the co-occurrence of the keywords within longer syntagmatic units, both at the phrase and the clausal levels. I used the Word Sketch functionality of Sketch Engine to calculate the LL and the LogDice measures.

2.2.2.1. Nouns

The first group of tags comprise NN2 (469.45) and NN1 (151.91). Plural nouns (NN2) are significantly more frequent in HEGP than in the BNCWS. These nouns include in decreasing frequency (>25 occurrences) *students, providers, institutions, reasons, groups, proposals, powers, outcomes, backgrounds, employers, metrics, skills, benefits, universities, courses, incentives, bodies, views, changes, levels, costs and fees*. The plural *students* alone accounts for 0.93% of all the words in the HEGP, while the singular form accounts for 0.44%.

The lemma *student* is the most significant item of this group. It is modified by *prospective* (11.36), *disadvantaged* (11.09), *BME* (Black and minority ethnic) (10.82) or *full-time* (10.52). “White students” (10.08) does occur in the paper, always in the context of BME students. *Students* tend to premodify *support* (11.51) or *protection* (11.18) as in

1. if accompanied by some form of guarantee of **[student]** protection as a condition of designation, both financial and [...].

and are the objects of verbs such as *help* (10.89), *protect* (10.87) or *support* (10.82) such as in

2. This forms a key part of our aim that TEF supports **[students]** from disadvantaged backgrounds.

Students in the HEGP *receive* (11.03):

3. All **[students]** receive effective support in order to achieve their educational and professional goals and potential.

or *study* (11.03) such as in

4. [...] where the majority of full-time **[students]** are studying for qualifications validated by organisations with UK DAPs will be offered the choice to [...].

Students appear in coordinated phrases with *employers* (12.28),

5. We want to see greater assurance for **[students]** and employers that the class of degree awarded by higher education providers is consistent across the sector.

taxpayers (11.78):

6. the OfS would protect the interest of **[students]** and taxpayers by using compliance powers.

and *stakeholders* (10.19):

7. The next levels up will allow differentiation to help **[students]** and others stakeholders compare courses.

A tendency is observed to portray *students* in binomial structures where they lie side by side with *employers*, *stakeholders* and *taxpayers*, arguably the type of lexis which is more likely to be found in business contexts. The second most prominent lemma in this category is *provider* (382), premodified by *alternative* (12.19), *new* (11.60), *education* (11.15), *quality* (10.37) and *publicly-funded* (9.63):

8. 77% of students receive the full maintenance grant (payable to those with a residual household income of up to £ 25,000), compared to 39% at publicly-funded **[providers]**.

In all the instances where the expression *publicly-funded providers* is used, a comparison is established with *alternative providers*. In particular, it is notable that HEGP includes reports from the Student Loans Company that show that “the profile of the student population at the group of alternative providers currently designated for student support is different to that at publicly-funded providers”. These students tend to be older, male, from low-income households and from a non-white ethnic group. Providers are the objects of *allow* (11.74):

9. It also sets out how we propose to deliver on the commitment, announced in the Productivity Plan, to allow **[providers]** with high quality teaching to increase their fees in line with inflation from the 2017/18 academic year.

and *require* (11.42):

10. A duty to operate a single-entry route into the higher education system, with powers to require **[providers]** to meet conditions attached to the model of entry chosen [...].

They are subjects of *deliver* (10.57):

11. We therefore propose some specific incentives for alternative **[providers]** delivering HE provision at level 6 under the TEF.

and appear juxtaposed with *disciplines* (10.97) as in one of the consultation questions:

12. Do you agree that the ambition for TEF should be that it is open to all HE **[providers]**, all disciplines, all modes of delivery and all levels?

Given the collocation profile of providers in the HEGP, it is hardly surprising that the answer expected will be affirmative. This consultation question is an example of how previous context can frame the answer in ways that favour the policy in a green paper. *Institutions* occurs with *administrative* (10.791), mostly in the questions section of the HEGP:

13. How can we minimise any administrative burdens on **[institutions]**? Please provide any evidence relating to the potential administrative costs and benefits to Institutions of the proposals set out in this document.

and with *burdens* (10.232) in the same context. *Outcomes* collocates with *gain* (11.439) in the string *student outcomes and learning gain*, with *environment* in *learning environment*, and with *employment* (10.631) as in:

14. A key focus of TEF should be the educational and employment **[outcomes]** of higher education, and the gains made by students from different backgrounds.

2.2.2.2. *The expression of finality*

The second group of tags comprise IF (234.73), TO (196.58) and BCL21 (56.56). The latter identifies *in order to* and it occurs 31 times in the document. It collocates with *maintain* (11.608), *achieve* (11.159), *make* (10.376) and *level* (9.857). The IF tag identifies uses of the preposition *for*, which appears after *reasons* mainly in the consultation questions (“please give *reasons* for your answers”), an overarching genre feature of green papers. *Students* (10.71) is significant as prepositional complement such as in:

15. The regulation of higher education must evolve to champion value for money **[for]** students making big lifetime investments,

as well as in the expression *value* (10.873) *for money* (10.834). The *Office* (10.778) *for students* is also a prominent occurrence in the text together with *education*, both at clausal level:

16. Widening participation in higher education is a priority **[for]** this Government and will help to drive social mobility.

and at phrasal level:

17. Degree inflation carries significant reputational risks **[for]** UK higher education: employers face a challenge distinguishing between graduates.

The TO tag identifies uses of the infinitive marker. Most of the verbs used after the tag are found in the context of a very restricted set of subjects in the infinitive construction such as *providers* (11.674) or objects such as *students*:

18. in order [to] protect students, ensure value for money for the public purse, and focus oversight where it is needed most.

Figure 1 shows a selection of concordance lines where, in most of them, the expression of finality is found together with verbs that convey the idea of giving, helping and providing avenues for students' participation.

risk-based approach to regulation, in order	<to > protect students , ensure value for money
propose to transform the regulatory landscape	<to > put students at its heart and create a
and where possible stimulate the sector	<to > help students meet their aspirations by
recognise those institutions that do the most	<to > welcome students from a range of disadvantaged
employers and the taxpayer. The aim is	<to > improve the teaching that students receive
recognise those institutions that do the most	<to > welcome and support students from a range
, but robust in its judgements in order	<to > maintain the confidence of students , the
as to which TEF level to award. In order	<to > provide useful information to students
next levels up will allow differentiation	<to > help students and others stakeholders compare
gain 9. Excellent teaching has the ability	<to > transform the lives of students . A key
students participating in the fund being used	<to > provide finance to future students who
increases choice for students , and can help	<to > deliver better value for money. Our aspiration
introduced by the last Government in order	<to > maintain quality, protect students and
can apply for specific course designation	<to > enable its students to access student support
course designation to enable its students	<to > access student support. Academic performance
higher education providers, making it harder	<to > recruit the best students . . Monetary
the highest quality alternative providers	<to > recruit more students by introducing a
value proposition to students will be able	<to > increase their share of total students
safeguard the ability for all students	<to > benefit higher education, maintaining a
providers to have contingency arrangements	<to > support students in the event that their
interests of students , employers and taxpayers	<to > ensure value for their investment in education
conditions can easily enter the sector	<to > give students more choice, while safeguarding
the onus on, and progress of, providers	<to > widen access for disadvantaged students
study . Ensuring the rights of students	<to > hold providers to account and get value
the rights of students to hold providers	<to > account and get value for money for their
following statutory duties and powers: . A duty	<to > promote the interests of students to ensure
to complete access agreements in order	<to > charge students fees above the basic amount
of the Director of Fair Access. . A duty	<to > provide students with accessible information
students. . A power to require providers	<to > meet conditions to protect students in
to require providers to meet conditions	<to > protect students in the event of course
providers to have contingency arrangements	<to > support students in the event that their
Excellence Framework 4. The aim of the TEF is	<to > help all students understand the quality
their teaching and grow their capacity	<to > teach more students ;.Changes in provider
withdraw. 19. The overall effect will be	<to > benefit all students of all backgrounds
providers to have contingency arrangements	<to > support students in the event that their

Figure 1. Concordance lines showing the expression of finality and students' participation.

These meanings contrast with those associated with *providers* in Figure 2, which are related to rules and norms, obligations and government supervision. This POS keyword analysis shows evidence that the occurrences of *students* and *providers* in the HEGP are distinctively primed to construct a view of the Government as guarantor of the rights of citizens and students. The Administration in

the HEGP will *ensure* (9.12) *value* (11.04) for money for taxpayers and will *allow* (11.74) providers to achieve (9.91) quality standards for expansion and, among others, fee increase to take place. The Administration will similarly *require* (11.42) these providers to meet conditions and base levels of quality and will monitor (9.91) them. However, there is room for the recognition that “the income of nearly all of these providers is no longer principally from direct grant”, which makes the case for *treating* (9.87) “nearly all” HE providers as non-public bodies:

19. Alternative providers are not **[treated]** as public bodies.

<p>job. 9. Higher education providers need employers as to which providers they can trust commitment, announced in the Productivity Plan, allow providers with high quality teaching at the moment new providers find it hard reforms gave providers the opportunity for improvement. More needs to be done quality of teaching and allow providers level of skills and harder for providers skills and harder for providers to know how backgrounds, 21. In allowing providers successful in TEF and providers will be entitled are sufficiently robust and can be used be pre-conditions. In particular we want there are incentives for all providers mean all eligible providers would be able transparent and fair way across all providers across all providers to give information additional evidence providers might want increased responsibility placed on providers Agreements for 2016/17 with plans for providers recognise the efforts that providers make new Office for Students to have the power Office for Students should have the power set targets where providers are failing through which all providers are equally able clearer and faster trajectory for providers example by allowing small and new providers elements of application processes, in order degree level providers would also be able This has led to many providers needing all predominantly degree-level providers placed on them, for example not being able university title and HEFCE designation, in order into the sector, providers will be able first time, and for the best providers will be some new providers that started quality assurance, allowing providers Governance and Management (FSMG) checks is 31. For alternative providers, in order education providers, making it harder little incentive for new providers that wish clear route available for these providers highest quality alternative providers performing providers were also eligible significant numbers of alternative providers unnecessary barriers to exit. 8. The proposal is to do so, for struggling providers so as</p>	<p><to > provide degrees with lasting value to their <to > produce highly skilled graduates. 22. Part <to > allow providers with high quality teaching <to > increase their fees in line with inflation <to > engage on a level playing field with the <to > grow and there have been signs that students <to > ensure that providers offering the highest <to > expand so that they can teach more students <to > know how to develop and improve their courses <to > develop and improve their courses. For <to > raise fees in line with inflation, we must <to > raise their fees in line with inflation <to > compare providers. Degree classification <to > ensure that providers applying for TEF <to > drive up standards. In our current regulatory <to > access the fee loan cap uplift and this <to > give information to inform assessments. <to > inform assessments. But we recognise that <to > offer and will consult further in the technical <to > widen access. As part of this framework <to > spend more than 745 million on measures <to > improve the access and experience of students <to > set targets for providers that are failing <to > set targets where providers are failing <to > make progress? Please give reasons for <to > select an operating model which works for <to > award their own degrees, and to secure <to > enter the sector with an annual student <to > simplify the framework for providers 7. <to > choose between which of the following approaches <to > demonstrate no fewer than four years experience <to > make a case for DAPs, and the Office for <to > validate degrees at other providers, continued <to > allow those providers who are ready to <to > apply for university title as part of their <to > grow, whilst establishing how best we might <to > deliver courses in September 2014 that <to > apply for HER after having applied for <to > ensure that providers with specifically <to > be able to demonstrate sufficient track <to > recruit the best students. Monetary and <to > enter the sector for the first time, as <to > grow their student numbers over time. We <to > recruit more students by introducing a <to > receive an increase of a further 10% (in <to > expand their student cohort and compete <to > introduce a requirement for providers to <to > ensure that England has a world class higher</p>
--	--

Figure 2. Concordance lines showing the expression of finality and HE providers.

2.2.2.3. Modality

The third group of tags comprises modal verbs (VM, 154.59). VM are significantly more frequent in HEGP (2.4% of all words in the text) than in the BNCWS (1.4% of all verbs in the corpus). The decreasing order of normalized frequency of these verbs is as follows: *would* (6.76/1000 words), *will* (6.34/1000 words), *should* (2.96/1000 words), *can* (2.77/1000 words), *might* (1.17/1000 words), *may* (1.02/1000 words) and *must* (0.84/ 1000 words). *Would* is the most significant modal verb in the list and is “extremely common in conversation and fiction” (Biber *et al.* 1999: 488). Figure 3 shows the most significant collocates for *would* in HEGP. We can appreciate how four words collocate significantly with this modal verb: *we*, the *Office for Student (OFS)*, *providers* and *change*. LogDice measures are provided in brackets clockwise in decreasing order of significance.

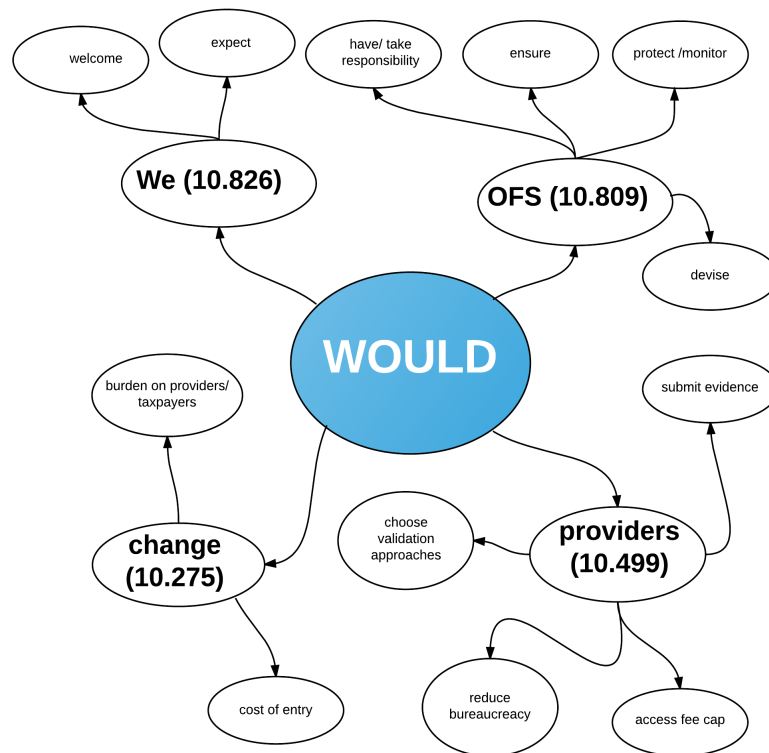


Figure 3. Significant collocates of *would*.

According to Biber *et al.* (1999) *would* and *will* are used to express prediction and volition. In conversation, these uses become blurred. However, in academic prose *would* is widely used for the prediction of events that do not involve any personal agency. In HPGP the use of *we* (PPIS2 tag) seems to be indeed very tightly connected with the expression of volition: *we will establish a regulator, we will reward excellent teaching, we want to maintain and even improve, we propose to transform the regulatory landscape*, etc. This use of *we*, however, is interspersed with uses where a more assertive voice emerges: *we must do more to ensure they (students) can make well-informed choices, we recognise that HE is not the only option for young people, we have the best universities, we believe that anyone with the talent and potential should be able to benefit from higher education*, etc. The combination of *we + would + verb* occurs 29 times in the text, typically to invite stakeholders to provide their views and opinions:

20. we **[would]** be interested in views as to whether the OfS should be able to contract out some or all of these functions in the way that HEFCE currently does.
21. But we **[would]** welcome your views on the broad principles outlined in this chapter.
22. We **[would]** welcome additional evidence from respondents to develop the evidence base further.

This invitation to participation is evidently inherent in the nature of green papers. What is required by the government here is views on the broad principles and additional *evidence*. This request is interesting as it positions the HEGP itself as evidence. One may wonder whether these principles, framed by the lexical choice discussed in 2.2.2.1 and, in particular, by the business drive identified as salient in the POS keyword analysis, will be perceived by the citizens differently. Similarly, asking for and receiving further evidence can most likely result in endorsing the policies advocated in the HEGP. When *would* is not used in this fashion, it is the Government voice that emerges more distinctively:

23. If there were separate bodies we **[would]** expect much closer strategic and operational co-operation between them.
24. We **[would]** expect the financial incentives to apply at the institutional level from the academic year 2018/19 and be differentiated according to the TEF level awarded;
25. Over the longer term we **[would]** like to be more ambitious. When the apprenticeship levy is introduced we **[would]** expect more employers to take advantage of excellent apprenticeship training offered by higher education providers.

We will occurs slightly more often in the HEGP, 35 times, and collocates with *consult* (11.964), *consider* (11.439), *look* (11.093) and *approach* (10.318). The Government here combines institutional agency and openness to further opinions and improvements in the TEF. The *OFS* is commissioned extraordinary supervision powers and is seen as the administration's driving force behind the intended changes. When these changes are made explicit, it is to note that taxpayers and providers should be relieved from further *burdens*. Providers are expected to reduce *bureaucracy*, submit *evidence* and, should they receive the go-ahead from the *OFS*, access the *fee cap* which may secure appropriate funding.

Should is the third most significant modal verb in the list and is "extremely common in conversation and fiction" (Biber *et al.* 1999: 488). Figure 4 shows the most significant collocates for *should* in HEGP. We can appreciate how six words collocate significantly with this modal verb: *TEF*, *open*, the *Office for Student (OFS)*, *measures*, *incentives* and *ambition*. LogDice measures are provided in brackets clockwise in decreasing order of significance. The uses of *should* in the HEGG reveal that, for the Government, the TEF is an opportunity to influence HE providers and either reward or refocus the performance of universities and similar bodies. Although it is not stated plainly and explicitly, the UK Government guarantees itself the power to streamline university funding. Contrarily, the *OFS*, the *measures* and the *incentives* in Figure 4 are only found in the question

sections of the document and are part of the consultation element in the HEGP.

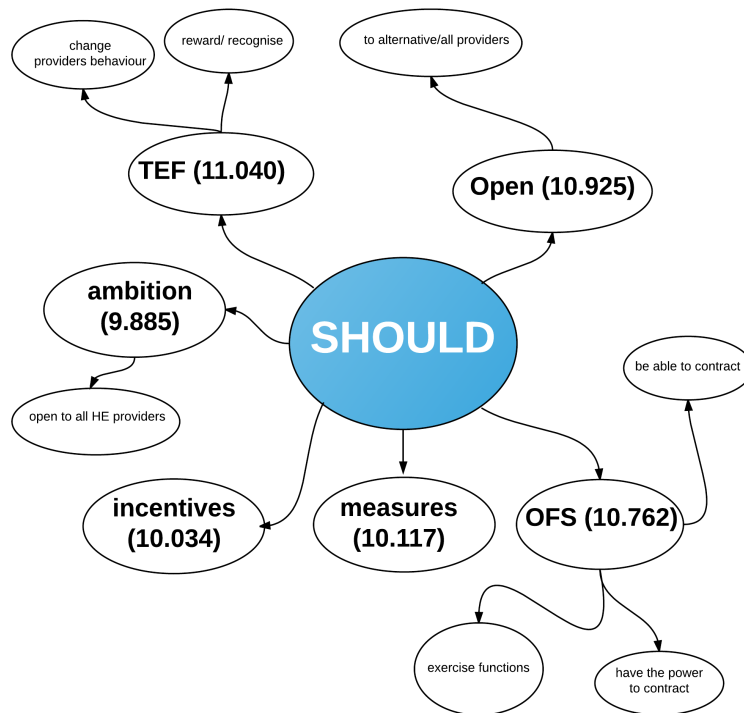


Figure 4. Significant collocates of *should*.

3. The Twitter debate on HEGP

3.1. Method

Zubiaga *et al.* (2011) used 7-day periods to monitor Twitter activity. In my research, two Twitter datasets were collected at two different

points in time: between 20 and 31 March, and between 17 and 28 November 2016. The first dataset was collected once the consultation period of HEGP had finished in January, while the second was collected well into the new 2016/2017 academic year and it featured the 19/11 United for Education demo in London called by the National Union of Students (NUS) of the UK. In this demonstration, the NUS voiced their criticism against rising tuition fees and, according to the organizers⁷, against a government that is forcing universities to run like businesses. The organizers encouraged the use of the #Nov19 hashtag on social media. Figure 5 captures a frame of an embedded video on a tweet that self-explains the reasons behind the protest:



7 URL:<<https://www.nus.org.uk/en/take-action/education/united-for-education>>

Figure 5. Tweet reporting the 19 November 2016 demo against the TEF.

Netlytic⁸ was used for the collection and preliminary analysis of these datasets for a variety of reasons. First, it offers automatic *keyword* extraction and network metrics; second, it lets users visualize the use of these words over time, and, thirdly, the service is provided free of charge and no programming skills are necessary to make sense of the data. After linking my own twitter account, I created a search query that contained the following terms: #TEF teaching excellence framework. #Nov19 was not included in the search to avoid bias. Netlytic scanned the dataset to find the most frequently used words and tallied them up. As most social media scanners, Netlytic ignores function words such as articles and prepositions. What the Netlytic developers labelled here as *keywords* is completely different from the analysis in the previous section. While Netlytic looks at the raw frequency of words, the POS keyword analysis in Section 2 adopted a more complex approach to lexico-grammatical saliency which involved the comparison against a reference corpus of British written English. When relevant, linked media will be discussed along with the keywords in the tweets.

3.2. Keywords in the analysed tweets

Table 2 shows the ten most frequent lexical items in the two Twitter datasets:

Dataset 1: March 2016		Dataset 2: November 2016	
Item	Instances	Item	Instances

8 URL: <<http://www.netlytic.org>>

Learning	148	#tef	42
Analytics	144	Teaching	18
National	132	@lseimpactblog	13
Service	130	#highered	12
Feed	103	Mock	11
@timeshighered	90	Results	11
Students	25	Hierarchy	10
Compare	23	Revealed	10
@chavergalthe	22	excellence	9
progress	22	@timeshighered	9

Table 2. Most frequent lexical items in the two Twitter datasets.

The lexical items in both datasets were totally different except for @timeshighered, the official Twitter account of Times Higher Education (THE) World University Rankings⁹. In the first dataset, 86 unique users retweeted an article published by Chris Havergal from THE on March 24: ‘National learning analytics service: could it feed into the TEF?’. In this article, the author explained how the UK would become the first country in the world to get a national learning analytics service to identify students who may be struggling and understand which teaching methods work best. The title of this article explains alone the lexical impact of the most frequent words in the March dataset. Other 21 users retweeted the same article, but, this time, these tweets put emphasis on an app that may allow students to compare their progress.

In the March dataset, the top ten most frequent words failed to convey any opinion whatsoever on the HEGP. The use of *view* is of interest. It was used eleven times to retweet the opinion of Professor Frank Coton, Vice-Principal (Academic and Educational Innovation), University of Glasgow. Essentially, he expressed concern over the compatibility of the TEF with the Scottish Quality Enhancement Framework (QEF), and stressed that, given the visibility of the UK universities in the new rankings, the universities in Scotland are committed to “engaging actively and constructively in the development of the TEF to secure a route to accreditation at all levels

9 URL: <<http://www.timeshighereducation.com>>

that is different but equivalent and builds on our existing, distinctive quality arrangements”. *Interesting* was used six times to introduce the retweet of the already mentioned THE article, while *insightful* is used three times to highlight the opinion by Bart Rienties, The Open University Reader in learning analytics, who said he would be sceptical of using the data in the TEF: “You have to wonder about the potential risks, because you have to understand the context that the students are working in.”

The range of frequent words that was generated by Netlytic and similar services needs to be taken cautiously and carefully examined against their co-texts. *Minors* is a good example. In our dataset, *minors* occurred four times and it was always a surname, not a common name. Kingston University Dr. Julia Minor’s participation in a roundtable discussion at the University of Bangor was retweeted four times. During her contribution, she was critical of the aims of the TEF although this aspect was not highlighted in any of the four tweets. Once one visits the transcript of her speech, her critical positioning emerges: “[the Government’s] trying to add to their leverage over how universities function, particularly how the teaching of our students is going to occur, and how that is valued, and what form of teaching or even assessment is going to be seen as acceptable”. Similarly, it is necessary to highlight how some of the most frequent lexical items in the March dataset are used to introduce minimal modifications to previous tweets or to introduce contents from other sources. *View*, *interesting* and *insightful* are used in this way.

The November Twitter dataset presents broader views on HEGP. *TEF* was used after a hashtag in 42 tweets. Many of those retweeted or reacted to an article (Figure 6) of the London School of Economics Impact of Social Sciences blog written by Marty Chamberlain, University of Southampton, where he discussed the role of peer observation of teaching in higher education in the context of the Teaching Excellence Framework (TEF).



Figure 6. RT of a LSE blog article.

The author remained sceptical of the benefits of peer reviewing of teaching in the current framework of UK universities:

We know that a significant amount of teaching is undertaken by staff on temporary and zero-hours contracts and promotions to senior posts are typically linked to research and funding capture, not recognition of teaching excellence. Will the TEF change this state of affairs? I doubt it. For me, a key question [...] is just how will the TEF change the way universities recognise and reward the range of talent and abilities their full-time and part-time staff possess, instead of continuing to reward just one part of the job – research – above all others?¹⁰

The author believed that the primary focus of TEF reviewing must be departmental and institutional performance, not individual performance and stated that the introduction of the TEF was arguably yet another example of the global trend towards neo-liberalization and marketization in HE. An article written by Anna Fazackerley for *The Guardian* – ‘Universities and NUS plan boycott of flagship teaching rankings’ – was very popular among the Twitter community in our

10 URL:<<http://blogs.lse.ac.uk/impactofsocialsciences/2015/11/23/peer-review-of-teaching-and-the-teaching-excellence-framework/>>

dataset. The word *boycott* sparked some interest and thus tweets like these were found:

26. Tough out #TEF? Unis and students may boycott 'teaching quality' ranking that doesn't include teaching.#TEFOff <https://t.co/DyjucNwH5X>
27. @NUSUK against #TEF & link to fee increase so may boycott the revamped #NSS.Performance tracking could be tricky!<https://t.co/7rh1TEwXix>
28. #Universities and #NUS plan to boycott new Teaching Excellence Framework rankings. <https://t.co/yIHub4M5Sp> #education #higher #TEF
29. Trouble with #TEF - #higher education boycott on rankings possible: <https://t.co/UHSGKNZ91w>

The above show some form of debate beyond retweeting as the users reshuffled words and concepts in an effort to accentuate their opposition to the TEF. Besides, there is a clear effort to build on hashtags that can provide the Twitter community with further visibility and greater potential impact. The NUS are specifically active in this area. Figure 7 shows a tweet where, both in terms of the language and the embedded graphics, straightforward opposition to TEF is visible.

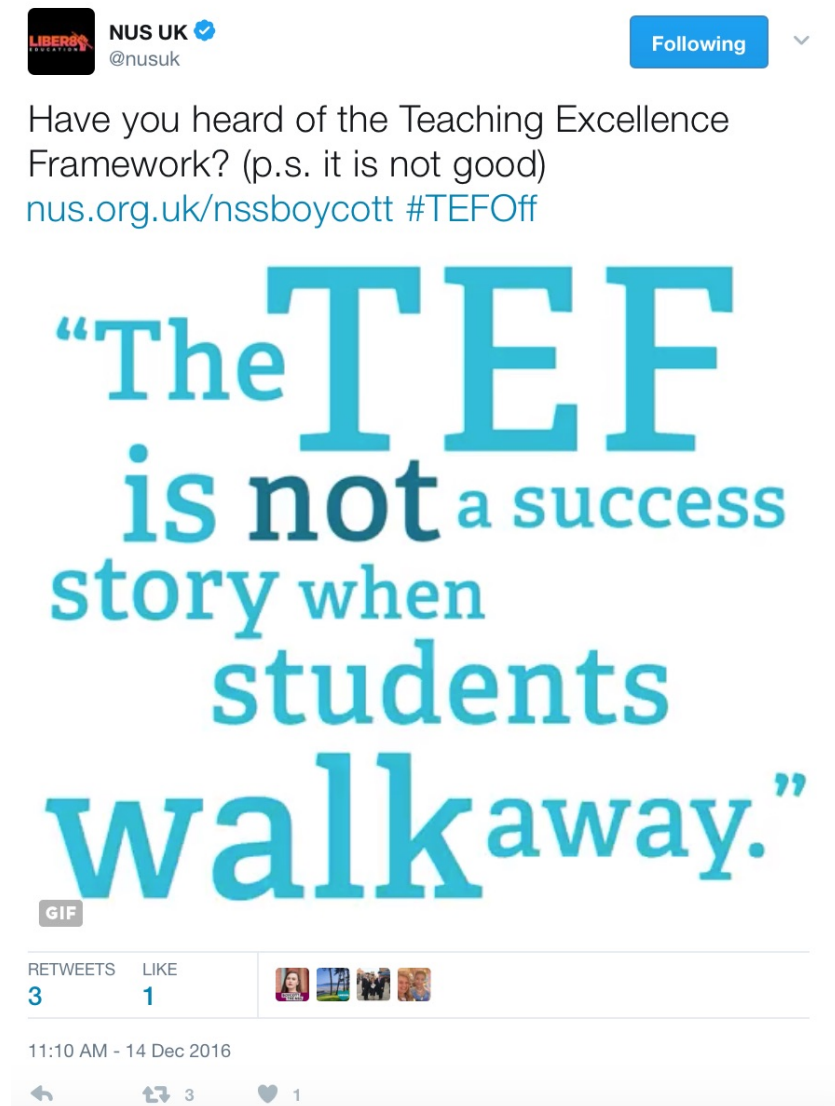


Figure 7. NUS tweet on the TEF.

As with the March Twitter dataset, the most frequent words are not particularly useful to gauge the personal views and opinions of the tweeters. *Teaching*, for example, is frequent in the November dataset

but it is typically linked to media such as Jessica Patterson's article¹¹ 'Five reasons the Teaching Excellence Framework is bad news for higher education', published in the London School of Economics Impact of Social Sciences blog. In her article, the author stated that the TEF meant moving more quickly towards the marketization of higher education and a 'value for money' approach to education. However, the tweets which contained the term *teaching* were mostly neutral and showed very limited involvement. An exception is a tweet where a well-known UK lecturer posted a picture of an apparently exhausted dog together with the tags #TEF #academia #phdchat #highered #AcademicsWithDogs to show her reaction to the TEF. *TEF* was mainly used in retweets of a THE article which included findings from mock TEF results that offered a totally different picture from the usual University Rankings, and which disfavoured Russell Groups Universities that no longer topped the new ranking. Words like *protest* and *students*, used six times in the November dataset, were always found together in the same retweet. *Government* and *system* were used five times each in exactly the same way. An example is #TEFinitelyNot, a hashtag widely used by the University of Manchester Student Union to voice the lack of suitability of the TEF for measuring excellent teaching. Even after considering hapax legomena, no traces of debate or discussion other than the ones presented above were found in the November dataset.

The two datasets are dissimilar in other ways. Netlytic also provides measures that capture how networks are organized. I looked at the networks in both datasets separately and examined centralization and modularity in both datasets. Centralization (Scott 2012) is a measure of the average degree of centrality of all nodes (the Twitter user accounts) within a network of users. Centralization describes the extent to which network graph cohesion is organized around some focal points. The March database yielded a centralization measure of 0.28 while the November dataset showed a centralization measure of 0.12, which suggests that information flowed more freely between more participants in the latter as high centralization values

11 URL: <<http://blogs.lse.ac.uk/impactofsocialsciences/2015/09/10/five-reasons-the-teaching-excellence-framework-is-bad-news-for-higher-education>>

closer to 1 imply the existence of a few central participants who control the flow of information in the network. The November dataset yielded a modularity index of 0.73 while the March dataset yielded 0.6. Higher values of modularity indicate clear divisions between communities as represented by clusters in Netlytic, which confirms that the tweets tended to overlap more significantly in the March dataset.

4. Discussion

4.1. *The Higher Education Green Paper*

In this chapter, I adopted a macroscopic Type III research approach in the tradition of corpus linguistics (Rayson 2008: 520) by looking at variation within the HEGP and examining “how certain features or groups of features characterise a text” in systematic comparison with a reference corpus. The POS keyword analysis in this study revealed that the HEGP is structured around three distinct language features that contribute to the endorsement of a conservative, neoliberal policy that presents a view of HE that puts economy before other concerns. The first is the use of nouns, both singular and plural, statistically more frequent in the HEGP than in the average written text as represented by the BNCWS. A set of nouns is specifically salient: *students*, *providers* and *institutions*. Those students depicted in the document reflect an explicit interest in minorities and disadvantaged learners. They are described as groups that require both *support* and *protection*. *Providers* are represented as *new* and *alternative*, and very often a contrast is established in the HEGP between *publicly funded* and *alternative* providers. Finally, *institutions* in the HEGP tend to be providers associated with *burdens*, in particular, *administrative* ones.

These results show that, grammatically, the HEGP favours a phrasal over a clausal style. This is a typical feature of academic writing (Biber/Gray 2015) where information is packaged to present a

more static rendering of the people and the institutions involved in the HEGP. While plural common nouns in the BNCWS are 5.75 % of the total tokens, in the HEGP plural nouns represent 8.9 %. Singular common nouns represent 15.2 % of the tokens in the BNCWS and 18 % in the HEGP. This preference for common nouns contributes to a representation of experience and reality in terms of the ideational function (Trappes-Lomax 2004), which organizes the readers' understanding of the intended aims of the HEGP. Beard (2000) has argued that all political argument is ideological and that the naming labels given to the participants in discourse is part of how arguments are shaped. In this context, HEGP readers are presented with a referential specification mean higher than the average mean in British English written texts. Our study shows that common nouns alone accounted for almost 27 % of every word in the HEGP, whereas both common and proper nouns are responsible on average for 33% of all words in English news (Biber *et al.* 1999), that is, the register with the highest density of nouns in the English language. The fact that specification is constructed around a small number of significant nouns reinforces the notion that the HEGP *aboutness* is restricted to both *students* and HE *institutions*, which somehow tucks away the agent role of the Administration in shaping important changes in UK HE. This role is partially subsumed by the use of the pronoun *we* together with the use of the modal *would*, and the introduction of a new regulator, the *OFS*, which again collocates significantly with *would*. Meaning specification in the HEGP is built around a very limited set of nouns that set the scene for a construction of subjects and relations (Fairclough 2001) which prime the economic benefit and a market orientation in HE.

The second distinctive language feature revealed by the POS keyword analysis was the expression of finality, in particular to convey the idea of supporting *students'* participation. Ironically, this effort was, apparently, poignantly ignored in the Twitter debate discussed in the previous section. When finality is expressed in the context of *providers*, the Administration adopted here a supervisor role that, among other actions, would enable HE institutions to increase academic fees, the most controversial measure in the whole green paper. In the new HE and Research Bill, the government plans

to rate universities gold or silver in the TEF. These universities will be able to raise their fees in line with inflation from 2019 onwards. Bronze-scoring institutions, however, will be eligible for *only* half of that rise. The House of Lords rejected these plans in early March 2017¹², but the HE and Research Bill has a long way to go before it is passed. In any case, assessments of providers based on their performance on student satisfaction, retention and graduate employment, as well as further submissions, are under way.

Modality is the third linguistic feature that surfaced after the POS keyword analysis. It is somewhat unusual that modals, and principally central modals like the ones discussed here, are so frequent in written registers. In fact, modality is only present in 12.5% of verb phrases across major registers (Biber *et al.* 1999: 456). These authors found that modals and semi-modals are much more frequent in conversation than in written registers, which turns our text into an interesting piece of discourse where modality stands out as one of its most defining features or ‘language form’ (Fairclough 2001). In fact, the use of modals in the HEGP served different roles in the HEGP. They were used to hide personal agency and to project an open attitude towards dialogue and agreement. In this sense, modal verbs contributed to the expression of different levels of authority (Baker 2011). When *should* is examined, one finds that there is an unconcealed desire to open up to new providers and this is the manifest ambition of the new HE policy as specified in the HEGP. Similarly, the Government wants to provide the new OFS with enough power so as to exercise control over the entire HE system. Paradoxically, by using the modal *should*, the Government indicates its superior status over the *OFS*, which, in turn, is a mere instrument of the Government to regulate and supervise the new HE domain. The modal verbs in the HEGP are thus determined by the power structure of the speech act situation (Winter/Gårdenfors 1995) but, simultaneously, operate epistemically by shifting the agency for some of the proposed avenues for action in the HEGP. While the uses of *should* are more deontic, we cannot forget that the *OFS* is a new

12 URL:<<https://www.timeshighereducation.com/news/house-lords-rejects-plans-link-tef-results-tuition-fees>>

regulator introduced by the Government in the HEGP to secure supervision and executive powers over HE institutions. Deontic modality is, in this context, directed towards the activities of the Government while epistemic modality is used to present the Administration as an open and understanding interlocutor that will listen to what society has to say about HE.

The results of this study support both Fairclough's (2003) claim that Green Papers contribute to legitimizing policy and emphasizing consensus through generalizing away, and McGettigan's (2013: 185) view that UK HE is being constructed as a market:

Education is being re-engineered by stealth through a directed process of market construction, each move designed to protect the elite and expose the majority. [...] Existing quality assurance, which has its faults, is supplanted by 'value for money', a 'risk-based' system, and a regulator tasked with promoting competition.

4.2. The Twitter debate

The analysis of both datasets revealed that Twitter users did not seem to rely chiefly on this social network to convey their own personal opinions and attitudes to HEGP. First, the number of tweets on both datasets was low. Second, most of the tweets analysed were retweets of media contents. This was certainly the case for all of the tweets in the March dataset, which showed a 0.28 centralization measure suggesting that very few tweeters controlled most of the exchange of information. In the March dataset, Twitter was mainly used to redistribute media and more specifically content generated by THE. In the November dataset, both the centralization and the modularity measures suggest that more tweeters took part in the debate and that the information distributed was both more diverse and depended less on one single source.

The results in this chapter suggest that the use of keywords based on raw frequency proved of little use for our research purposes. Manifestly, numerous retweets of a few tweets may create the false impression that those taking part in the debate are using a set of

lexical items to convey their own opinions. Words like *learning* or *national* were found to be very frequent in the March dataset but they actually were just part of the same string that was retweeted over and over. This finding confirms the problems identified by Teh *et al.* (2015) that suggested that words alone are not enough to assess automatically sentiment in commercial websites. Similarly, I may suggest that higher modularity values may be a good indication that keywords may represent more personal language choices, as opposed to mere retweets. This will, however, need further analysis and empirical confirmation. These results suggest that it is in the very infrequent words or hashtags where opinions on the TEF could be found and, no doubt, all of them voice their concern over the mercantilist turn in UK HE. Admittedly, this criticism was found in the links provided in the tweet rather than in the tweets themselves. Nazir *et al.*'s (2016) suggested that 30 % of academic staff use social networking sites as part of their everyday academic life. If this is the case, certainly UK scholars restrained from expressing their opinions on Twitter at least in the two time frames when the datasets were collected.

5. Conclusion

The findings in this chapter are to be taken cautiously. First, the control corpus in the POS keyword analysis was a representative corpus of written British English. A control corpus of green papers would have most likely yielded different results. Future research efforts should consider gathering such a resource. Second, despite following the recommendations in Zubiaga *et al.* (2011), the Twitter datasets were collected during two relatively short time frames for a non-trending topic such as the TEF or the HEGP. Longer time frames would possibly have enriched this discussion. Thirdly, the methods described in this paper are complementary rather than excluding

others widely used in discourse analysis. These results could be thus complemented and enriched by means of other research methods.

The reduced group of nominals in the HEGP was integrated by nouns with no technical connotations and, in this sense, it may prove hard to describe the HEGP as an ‘opaque text’ (Orts 2015). However, I would argue that some sort of opaqueness is present in the way in which some core ideas are delivered through a dense process of domination through language (Trappes-Lomax 2004). Power distance was not achieved by the use of specialized language. Rather, it was achieved by way of repetition of some very strong collocates (LogDice measures provided in brackets) such as *alternative* (11.9) providers, *high quality* (10.63) providers or *open* (9.85) to all HE providers. In other words, the representation of providers is biased to present an idealized business model that delivers quality and excellence. This is assumed by the HEGP authors who placed great emphasis on removing barriers to completion in the best interest of students. This positive, neoliberal role of competition was contested by media and HE experts that, judging from the range of media sources in the Twitter datasets, rejected these claims. However, the use of finality and the role of modality contributed to the manipulation of the readership of the HEGP by constructing discourse (Van Dijk 2006) in ways which, apparently, gave the Government the agency to control HE providers while, in reality, they created the conditions for further marketization of UK universities (Van Dijk 2006).

The methods in this chapter constitute a lexical approach to discourse analysis. I presumed for the POS keyword analysis that meaning is constituted, revealed, and constrained by the collocation environment (Schroeter/Veniard 2016) and previous uses of language (Baker 2011). As Sinclair (1965:76) put it in one of his early works: “Any stretch of language has meaning only as a sample of an enormously large body of text; it represents the results of a complicated selection process, and each selection has meaning by virtue of all the other selections which might have been made, but have been rejected.” POS keyword analysis offers researchers a look into what was rejected by examining patterns of use in terms of frequency and distribution across linguistic registers. In doing so, we are better prepared to analyse what we have in front of us. As for the

Twitter database, our results seem to confirm that the UK academic staff was either largely uninterested in taking part in it or refrained from doing so. The tweets examined all reject the implementation of the TEF and the industrialization of HE in the UK. This rejection has had no effect. In 2017, year two of the implementation of the TEF, HE institutions will be rated Gold, Silver, or Bronze based on different metrics that will look at graduate employment, student retention, and student satisfaction as determined by the National Student Survey, as well as on a narrative. Outcomes in Year Two will not be associated with differential fee uplifts for providers in England – rather, all those achieving a rating of Bronze, Silver and Gold will receive the full inflationary uplift. However, these awards will be used from Year Three onwards to inform differentiated fees¹³. The *not-so-obvious* marketization discourse in the HEGP revealed through the use of POS keyword analysis in this Chapter is consistent with market-making strategies that require direct government intervention (McGettigan 2013). The situation is creating a general feeling of confusion that Collini (2012) has framed as a misleading analogy between a university and a commercial company, where terms such as *brand recognition*, *output*, *efficiency* and *reduced costs* coexist with more traditional HE values.

References

- Baker, Paul 2011. *Using Corpora in Discourse Analysis*. London: Continuum.
- Baker, Paul / Gabrielatos, Costas / Khosravini, Majid / Krzyżanowski, Michał / McEnery, Tony / Wodak, Ruth 2008. A Useful Methodological Synergy? Combining Critical Discourse Analysis and Corpus Linguistics to Examine

13 14URL: <http://www.hefce.ac.uk/media/HEFCE,2014/Content/Pubs/2016/201632/HEFCE2016_32.pdf>

- Discourses of Refugees and Asylum Seekers in the UK Press. *Discourse & Society* 19/3, 273-306.
- Baker, Paul / Gabrielatos, Costas / McEnery, Tony 2013. Sketching Muslims: A Corpus Driven Analysis of Representations around the Word 'Muslim' in the British Press 1998-2009. *Applied Linguistics* 34/3, 255-278.
- Beard, Adrian 2000. *The Language of Politics*. London: Routledge.
- Biber, Douglas / Johansson, Stig / Leech, Geoffrey / Conrad, Susan / Finegan, Edward 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Biber, Douglas / Gray, Bethany 2015. *Grammatical Complexity in Academic English: Linguistic Change in Writing*. Cambridge: Cambridge University Press.
- Collini, S. (2012). *What are universities for?* London: Penguin.
- Fairclough, Norman 1999. Global Capitalism and Critical Awareness of Language. *Language Awareness* 8/2, 61-73.
- Fairclough, Norman 2001. *Language and Power*. Harlow: Pearson Education.
- Fairclough, Norman 2003. *Analysing Discourse: Textual Analysis for Social Research*. London: Routledge.
- Jeffries, Leslie / Walker, Brian 2012. Keywords in the Press. A Critical Corpus-assisted Analysis of Ideology in the Blair Years (1998-2007). *English Text Construction* 5/2, 208-229.
- Kilgarriff Adam / Baisa, Vit / Bušta Jan 2014. The Sketch Engine: Ten Years On. *Lexicography* 1/1, 7-36.
- McGettigan, Andrew 2013. *The Great University Gamble*. London: Pluto.
- Nazir, Salsabeel / Makhdoomi, Ursil / Jahangir, Waseem 2016. Usage of Social Media as a Research Tool – A Study of Research Scholars of Kashmir Valley. URL: <https://www.researchgate.net/publication/299344382_Social_Media_and_Researchers>
- O'Halloran, Kieran 2010. How to Use Corpus Linguistics in the Study of Media Discourse. In Anne O'Keefe / Mike McCarthy (eds) *The Routledge Handbook of Corpus Linguistics*. London: Routledge, 563-577.
- Orts, María Ángeles 2015. Power and Complexity in Legal Genres: Unveiling Insurance Policies and Arbitration Rules. *Inter-*

- national Journal for the Semiotics of Law – Revue internationale de Sémiotique juridique* 28/3, 485-505.
- Pace-Sigge, Michael 2013. *Lexical Priming in Spoken English Usage*. New York: Palgrave MacMillan.
- Pérez-Paredes, Pascual / Aguado, Pilar / Sánchez, Purificación 2017. Constructing Immigrants in UK Legislation and Administration Informative Texts: A Corpus-driven Study (2007-2011). *Discourse & Society* 28/1, 81-103.
- Rayson, Paul 2008. From Key Words to Key Semantic Domains. *International Journal of Corpus Linguistics* 13/4, 519-549.
- Schroeter, Melani / Marie Veniard 2016. Contrastive Analysis of Keywords in Discourses: *Intégration* and *Integration* in French and German Discourses about Migration. *International Journal of Language and Culture* 3/1, 1-33.
- Scott, John 2012. *Social Network Analysis*. London: Sage.
- Scott, Mike / Tribble, Christopher 2006. *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Amsterdam: John Benjamins.
- Sinclair, John 1965. When is a Poem like a Sunset? *A Review of English Literature* 6/2, 76-91.
- Sinclair, John 2004. *Trust the Text: Language, Corpus and Discourse*. London: Routledge.
- Stubbs, Michael 2007. On Texts, Corpora and Models of Language. In Stubbs, Michael / Hoey, Michael / Teubert, Wolfgang / Mahlberg, Michaela (eds) *Text, Discourse and Corpora: Theory and Analysis*. London: Continuum, 127-162.
- Taylor, Charlotte 2014. Investigating the Representation of Migrants in the UK and Italian Press: A Cross-linguistic Corpus-assisted Discourse Analysis. *International Journal of Corpus Linguistics* 19/3, 368-400.
- Taylor, Charlotte in press. Togetherness or Othering? *Community and Comunità* in the UK and Italian Press. In Chovanec, Jan / Molek-Kozakowska, Katarzyna (eds) *Representing the Other in European Media Discourses*. John Benjamins.
- Teh, Phoey Lee / Rayson, Paul / Pak, Irina / Piao, Scott 2015. Sentiment Analysis Tools Should Take Account of the Number of Exclamation Marks!!!. Paper presented at *The 17th*

- International Conference on Information Integration and Web-based Applications & Services (iiWAS2015)*. New York.
- Trappes-Lomax, Hugh 2004. Discourse Analysis. In Davies, Alan / Elder, Catherine (eds) *The Handbook of Applied Linguistics*. London: John Wiley & Sons, 133-164.
- Van Dijk, Teun 2006. Discourse and Manipulation. *Discourse & Society* 17/3, 359-383.
- Winter, Simon / Gärdenfors, Peter 1995. Linguistic Modality as Expressions of Social Power. *Nordic Journal of Linguistics* 18/2, 137-165.
- Zubiaga, Arkaitz / Spina, Damiano / Fresno, Victor / Martínez, Raquel 2011. Classifying Trending Topics: A Typology of Conversation Triggers on Twitter. In *Proceedings of the 20th ACM international conference on Information and knowledge management* New York: ACM, 2461-2464.