CARMEN DAYRELL
LANCASTER UNIVERSITY

CHAKRAVARTHI RAM-PRASAD
LANCASTER UNIVERSITY

GWEN GRIFFITH-DICKSON
THE LOKAHI FOUNDATION

BRINGING CORPUS LINGUISTICS INTO RELIGIOUS STUDIES: SELF-REPRESENTATION AMONGST VARIOUS IMMIGRANT COMMUNITIES WITH RELIGIOUS IDENTITY

ABSTRACT

This paper uses a type of social scientific data, obtained by semi-structured interviewing, to investigate the extent to which and the ways in which corpus linguistics methods contribute to research in the field of Religious Studies, and in the humanities and social sciences at large. The corpus consists of 73 interviews (357,788 words) with minority communities living in the UK, from various religious (Muslims, Hindus and Christians), ethnic and cultural backgrounds. All interviewees had a strong religious identity and were considered to be well-integrated into British society; either economically, socially, or both. These interviews were conducted in 2005 as part of a government-commissioned study broadly oriented to policy-makers, but located within the disciplinary approach of Religious Studies. Here, we examine how the various communities perceived the role played by their religious faith in the process of establishing themselves in Britain. The analysis seeks to identify common patterns of self-representation, that is, discourse patterns that contribute to the collective representation of each group, focussing on patterns cutting across groups. The major contribution of this paper is to assess how corpus linguistics methods can complement, refine and offer new insights to the type of discourse analysis currently established within the humanities. At the same time, we seek to test the limits of corpus methods, given the data might not be either qualitatively or quantitatively apt in all respects for corpus techniques.

CONTACT

Carmen Dayrell, ESRC Centre for Corpus Approaches to Social Science (CASS), Department of Linguistics and English Language, FASS Building, Lancaster University, Lancaster, Lancashire. LA1 4YW. c.dayrell@lancaster.ac.uk

ORCID

0000-0003-0165-0884, 0000-0003-4219-0415, 0000-0002-4402-6902

Bringing Corpus Linguistics into Religious Studies: Self-representation amongst various immigrant communities with religious identity

Carmen Dayrell
Lancaster University

Chakravarthi Ram-Prasad
Lancaster University

Gwen Griffith-Dickson
The Lokahi Foundation

1. Introduction

The combination of corpus linguistics with (critical) discourse analysis is a now-established field, but the use of corpus methods beyond the discipline of Linguistics is a more recent development (Ancarno, 2018). This paper explores one particular interaction between material gathered by experts in Religious Studies – for the purposes of exploring the sociology of immigrant religion in British society – and analysis undertaken through corpus linguistics techniques. We acknowledge that a different configuration of this interaction, as one directly between the discipline of Religious Studies and that of Corpus Linguistics, might have generated different conclusions.

The corpus consists of 73 interviews (357,788 words) with minority communities living in the UK, from various religious (Muslims, Hindus and Christians), ethnic and cultural backgrounds. These interviews were conducted in 2005 as part of a study commissioned from the Lokahi Foundation[1] by the Home Office, which sought to understand the impact of religious faiths on the integration of successful immigrant individuals into the British society. The two principal investigators of that project are two of the authors of the present study (Ram-Prasad and Griffith-Dickson). Section 3 details the context and aims of the original project; a full description of the data is provided in Section 4 and a brief summary of results in Section 5.

The data was collected to meet the specific needs of a government-commissioned study and followed certain conventional social-science methods, requirements and principles of the discipline of Religious Studies. This is precisely what makes this set of data ideal for the type of methodological discussion intended here. What is more, the corpus was *not* designed to be a representative sampling of the multicultural diversity inherent in religious immigrant communities living in the UK, but rather to create a specific selection about whom no claims of representativeness were made. It is thus not our intention to make generalisations about the discourse of UK's minority or immigrant communities in general.

Our primary purpose is to use a type of social scientific data to explore the extent to which and the ways in which corpus linguistics methods can be integrated into research in the field of Religious Studies, and the humanities and social sciences at large. The specific goal is to examine how relatively well-established sections of the UK's immigrant or minority communities perceived the role of their religious faith in

---

[1] The Lokahi Foundation is an independent research organisation (politically and religiously independent) which specialises in religious issues. See details at https://www.lokahi.org.uk/

the process of establishing themselves in Britain. Although individual experiences are brought into the discussion whenever relevant, the idea is to identify linguistic patterns contributing to the collective representation of each group. The analysis aims to identify discourse trends that cut across groups, rather than contrasting differences.

The major contribution of this paper is to reflect upon the difficulties and challenges inherent in the process of developing an interdisciplinary corpus analysis that meets the expectations and needs of another discipline. Our original data poses a particularly poignant question: the original study, in its methods and conceptual framing, presupposed that the uniqueness of each individual journey, the personal story-telling, was precisely the point. Any insights generated rested on highly interpretive, hermeneutical approaches of qualitative data analysis in order to generate meanings and insight. Could such contrasting methods generate results that are non-trivial for the other discipline? Can we map areas of utility but also areas of breakdown at the interface of the two disciplines? From the perspective of Religious Studies, there is the potential risk that using corpus techniques could yield the uninformative result that Christians speak about the Bible and the church, whereas Muslims speak about Islam and the mosque; and that our two humanities researchers would miss the richness of individual stories and deeper insights. On the other hand, the corpus approach ought to add empirical rigour to the analysis. By allowing for a systematic analysis of the discourse, corpus methods pinpoint prominent linguistic patterns irrespective of the researcher's intuitive prediction, thus reducing cognitive biases (Baker, 2006, pp.10–12). Our approach therefore enables us to assess how corpus methods can complement, refine and offer new insights to the type of discourse analysis currently established within the humanities, and Religion Studies in particular.

2.   The disciplinary context of this study

Whether undertaking a corpus-assisted discourse study (CADS, Partington *et al.* [2013, pp.10–14]) or using corpus methods in critical discourse analysis (CDA, Baker *et al.,* 2008), most studies that have combined corpus linguistics and discourse analysis have been carried out on large scale. The focus has been mainly on written language, such as the press (e.g. Baker *et al.,* 2008, 2013; Blinder and Allen, 2016), online textual material (Brookes and Baker, 2017), early English books (McEnery and Baker, 2016) or legal documents (Potts and Kjaer, 2015).

Closely related to this study is the interdisciplinary work by Nolte *et al.* (2018), which examined inter-religious relations among Muslims and Christians in the Yoruba-speaking south-west region of Nigeria, using a set of open questions asked in an ethnographic survey with 2,819 respondents (454,523 words). The data was collected as part of an anthropological project, entirely without the thought of a corpus analysis. Although their findings are not directly comparable with those of the present research, given the two contexts are fundamentally different, their overall conclusions are worth noting. While confirming existing anthropological literature related to the Yoruba context, the corpus investigation also uncovered discourse patterns that

traditional ethnographic methods alone would not have permitted. At the same time, anthropological knowledge of Yorubaland was essential to adequately interpret the results.

In terms of transcribed speech, most corpus studies have been based on corpora especially created for linguistic research, such as the British National Corpus (McEnery *et al.*, 2017) or the Louvain International Database of Spoken English Interlanguage (Gilquin *et al.*, 2010). A noteworthy exception is Sealey's (2009, 2012) analysis of discourse patterns in social behaviour, using a corpus of 144 oral history interviews (1.8 million words) recorded by historians to reflect social-cultural experiences of residents of the city of Birmingham (UK).

The present study also innovates by examining the intersection between religion and immigration across different immigrant communities. The vast majority of the literature on immigration has historically focused on race (Small & Solomos, 2006) or, more recently, on the study of the normative challenges of immigration in the context of British polity (Calder *et al.*, 2010; Webber, 2014). The literature also tended to locate immigration within debates about the nature of British society and culture (Ford, 2011; Kaufmann and Harris, 2015); although some recent studies have corrected these biases by examining immigrant responses to their place in Britain (Bauer, 2018). Research into the nuanced role of religion within British immigrant communities has mainly focused on a specific religious faith (Chivallon, 2001; Sharma, 2012), mostly Islam (Geaves, 2015; Lewis, 2007; Lewis and Kashyap, 2013; Maliepaard and Schacht, 2018; Werbner, 2002).

The picture is no different within the field of discourse analysis. Scholars have focused either on immigration or religious issues; research on the intersection between the two aspects is almost inexistent. Critical discourse analysts have concentrated on investigating discursive constructions of immigrants in relation to ideologies, power or racism (Capdevila and Callaghan, 2008; Van Dijk, 1991). Whether undertaking CDA or CADS, corpus linguists have focused on characterisation of immigrants in the British press (Baker *et al.*, 2008; Blinder and Allen, 2016) and, more recently, in official government reports (Pérez-Paredes *et al.*, 2017). As for religious faiths, scholars have focused on Islam specifically, uncovering evidence for either the negative representation of Muslims in the British press (see Baker *et al.* [2013, pp.17–19] for a review on the topic) or Islamic extremist discourse (Prentice *et al.*, 2012).

The present analysis differs from previous research in various ways. First, it looks across different groups coded by both race/ethnicity and religion. Second, it gives access to participants' views on their experiences and life in the context of immigration, being a member of a minority community, religious values and identity. Third, it brings out the socio-economic success that these individuals and communities had growing up as a minority in Britain, or moving to and establishing themselves in Britain, and the role their religion played in this.

3. The original research project

The original research project was carried out in 2005-2006. Prior to July 2005, the New Labour government was already responding to twin pressures relating to

multiculturalism in Britain: contentious political debates about immigration, and an increased interest in 'community cohesion' in the wake of the Oldham riots with the subsequent Cantle report and other policy responses. Thereafter in July 2005, the first Al Qa'ida-inspired terrorist attack on British soil occurred, and cohesion, religious-ethnic tensions, and public safety took on a new urgency. The government had no framing for successful models in cohesion and integration, and were receptive when the Lokahi Foundation proposed a study investigating various religious and ethnic minority community members who were 'success stories' by a variety of criteria: economic, social or artistic-cultural.

The strategic rationale behind the original project was that insights could be gained and lessons learned by listening to the voices from a variety of minority backgrounds, to help shape policy. The Lokahi Foundation was quietly aware of the fact that senior individuals who played a key role in commissioning, receiving, or ignoring the study, were divided on the tactical question of whether it was best to focus on Muslims as a distinct community requiring a policy response; or whether all interventions (and government funding) were best approached as a 'whole-community', or 'interfaith' response. Thus the deeper strategic philosophy, so to speak, was tacitly a critical question for policy, not only in community cohesion but also public safety and counter-terrorism. The research aim therefore was two-fold: (i) to investigate whether successful individuals from different religious and cultural backgrounds used broadly similar strategies and skills or whether their approaches differed according to their religious teachings or ethnic identity; and (ii) to identify any positive impact of religious faith on successful pathways of integration and understand how religious values interacted with other factors such as education, social status, family environment and personal qualities. In the simpler language of the consumers of the research, their urgent questions were: what 'works' in integration and community cohesion? And is religion only a problem in community tension or can it be part of the solution?

Aware of this ideological backdrop as well as the potential backlash against stigmatised communities, the project determined to learn from positive stories of success; and hear from those who were *not* seen as needing government attention – i.e., whose very success in integration had rendered them invisible to problem-oriented policies. Sites selected were in the North of Britain, in Lancaster and Preston; in the Midlands, in Birmingham; and in London. The groups as shown in Tables 1 and 2 were therefore: (i) African and Caribbean Christians in London (CR); (ii) South Asian and Middle Eastern Muslims in London (ML-SA); (iii) Gujarati Hindus in Preston (HD-GJ); (iv) South Indian Hindus in Birmingham (HD-SI); and (v) Gujarati Muslims in Lancaster (ML-GJ). Finally, the study included a group of particular high achievers, designated as the Super-Group (SG), who were considered a distinct and single sub-group, regardless of religious adherence or ethnic identity, whilst all being from an ethnic or religious minority. These consisted of extraordinary success stories, whether in financial terms, artistic or cultural high achievers, or other outstanding individuals in law, finance, or public service.

4. Data collection in the original study

Table 1 summarises the composition of the corpus in terms of religious groups, number of interviews, interviewees and words from each group. The data includes two focus groups and one interview with two female students.

Religious groups	Number of interviews	Number of interviewees	Number of words
Christians (CR)	12	12	64,650
Hindus (HD)	25	28	106,314
Muslims (ML)	29	33	141,481
Super Group (SG)	7	7	45,343
Total	**73**	**80**	**357,788**

Table 1: Composition of the corpus

The Hindu and Muslim groups comprised two distinct communities each (Table 2). The HD-SI and ML-SA subgroups included a focus group with four participants each.

Religious groups	Sub-group	Number of interviews	Number of interviewees	Number of words
Hindus (HD)	Hindus from Gujarat (HD-GJ)	13	13	43,528
	Hindus from Southern India (HD-SI)	12	15	62,786
Muslims	Muslims from Gujarat (ML-GJ)	18	22	77,403
	Muslims from Southern Asia (ML-SA)	11	11	64,078

Table 2: Composition of the Hindu and the Muslim subcorpora

All interviews were semi-structured, with a set schedule for the different interviewers to follow. However, it was important to conduct the interviews in a relaxed, informal, and conversational style. This was to ensure good rapport for methodological as well as ethical reasons. Further, given the likely diversity of the interviewees, interviewers were trained to follow the lead of the participant and their story, and thus could follow up, follow a direction indicated by a participant, or abandon a line of questioning if it proved to be irrelevant or inappropriate for that participant. The questions covered demographic information, religious practice and salience, then asked about their self-understanding of successes and failures, diversity of their friendship groups and contacts, and their self-understanding as part of a British community.

The interviewers included two professorial principal investigators (one Indian Hindu, one white-other Christian, one male and one female), research associates of the Lokahi Foundation, and postgraduate students taking a Master's degree in Religious Studies at Lancaster University at the time. They were from various ethnical backgrounds: two white British women, one British Egyptian Muslim woman, one British Caribbean woman, and one British Pakistani man.

Most interviewees were born somewhere other than Britain (64%). Within these, the majority (69%) came to the UK in the 1950-60s, at an early age; some had been living in the country for over 30 years. Only five interviewees had been in the UK for less than 10 years; they were all Hindus from India, and either highly educated or semi-skilled professionals. British-born, thus second or third generations of immigrants, accounted for 36% of interviewees; in most cases, either both or one parent came to the UK at an early age.

All interviewees were economically integrated; they were either gainfully employed, ran their own business successfully, or were students. They worked in a wide range of professional fields: accountants, medical doctors or nurses, school teachers, solicitors, university lecturers or entrepreneurs. The ML-GJ differed from the other groups with respect to qualifications and occupations. Only one out of the 22 interviewees held a university degree; most did not go further than secondary school, a few completed a college qualification. Their occupations varied greatly: students, housewives, restauranteurs, waiters and teachers.

Most interviewees were married, usually to someone from similar background. In terms of gender, there was an overall balance between female and male. As for age range, Christians, Hindus and the Super-Group were mostly mature adults, ranging from early 30s to 75. The age range among Muslims was lower; many were under 30.

Interviews were transcribed and each file was assigned a unique identification. The transcribed texts were marked up for speaker turns, considering the interviewers and interviewees. This annotation allowed us to isolate the interviewees' voices, which is the focus of the study. The number of words above (Tables 1 and 2) therefore refers to the amount uttered by the interviewees solely.

5. Data analysis in the original study

The original study made use of a method of qualitative data analysis constructed by the principal investigators (Griffith-Dickson and Ram-Prasad), which was hermeneutical and adapted to the multicultural, multi-religious material. It combined the use of researcher-generated themes and codes, applied to all samples, along with themes and codes arising from individual participants to generate understanding of each individual journey using the participants' own metaphors, concepts, and verbal tags. Each interview was thus interpreted uniquely, in a manner similar to Interpretative Phenomenological Analysis (Smith, 2009); subsequently, themes and insights were compared and aggregated across the piece. Finally, these themes and insights were translated into 'policy-maker-friendly' observations, to later generate recommendations.

The results were not put into public domain but a brief summary is given here to illustrate the nature of the material generated. The first, unexpected finding was that on all the observations below, there was no discernible difference across the sample – not between religions or ethnic origin. The attitudes recorded here occurred in all the groups. The most striking finding was that all, regardless of religious background, felt that their religion gave them an obligation to help others, and in particular to give something back to the community. Religion was a strong motivator to succeed, gave them support in difficult times, helped them overcome difficulties including racism and be a change-maker, and enabled them to understand their challenges within a larger interpretive framework. They identified with the label British if they could express it in a way that acknowledged their intersectionality, and contribute to what 'British values' are understood to consist in.

6. Methodology of the present study

The analysis builds on corpus linguistic techniques typically used to identify recurrent linguistic patterns in the discourse: *keywords* and *key semantic domains* (semantically-related words, [Rayson, 2008]). For these calculations, we used the interviewees' discourse only, thus discarding the language produced by the interviewers.

We first employed the techniques to identify the most salient words and semantic domains within each group by comparing each set of data with a one-million-word sample of the spoken component of the 1994 British National Corpus (BNC)[2]. Using the software tool *WMatrix* (Rayson, 2009), we combined a statistical test of significance (log-likelihood) with an effect-size measure (Log-Ratio — [Hardie, *forthcoming*]). For the former, we established the minimum critical value of 6.63 ($p <$ 0.01). Log-Ratio (LR) was used to rank the keywords and semantic domains in decreasing order so that the most salient items figured at the top, discarding those with LR lower than 1.5. The frequency thresholds were 30 and 50 occurrences in the study corpus for keywords and semantic domains respectively. To increase the likelihood of an adequate semantic categorisation, we set 'Religion and the Supernatural' and 'Kin' as preferred domains. The latter was necessary because the words *father* and *brother* referred mostly to family members, rather than members of the clergy.

As there were fundamental differences between the selected groups in terms of religious beliefs, ancestral lands, practices and ways of life, the next step was to identify differences between groups by calculating keywords and semantic domains for each group versus a combination of the others. We first focused on the CR, HD and ML and compared each with the other two combined. The SG was discarded from this calculation because it comprised a mix of religions and backgrounds which, at least in principle, could neutralize differences between groups. We then compared the SG with the other three combined (CR + HD + ML). For these calculations, we adopted the same statistical measures and thresholds as described above.

---

2   See http://www.natcorp.ox.ac.uk/; accessed on 13/03/2020. Importantly, we also applied the techniques using the spoken BNC2014 as a reference corpus (http://corpora.lancs.ac.uk/bnc2014/; accessed on 13/03/2020). No significant differences were found between the two versions.

The analysis focused on the 10 most salient (highest LR) keywords and the five most salient semantic domains in each subcorpus. This was because we intended to retrieve similar amount of information from all datasets given that our goal was to identify the most salient discourse patterns within each group, rather than contrasting differences across groups. These thresholds were nevertheless used as the starting point for the analysis and we considered items below the thresholds whenever necessary.

This initial quantitative analysis provided a broad overview of the most dominant discourse patterns and pointed towards words that merited closer investigation. Using *LancsBox* (Brezina *et al.*, 2015), we analysed all instances of the selected words through close reading of their concordance lines, going as far back or forward in the text as it deemed necessary. We also considered the distribution of patterns across texts so as to avoid concentrating on features that were restricted to a handful of texts.

7. Data analysis

This section starts with an overview of the main topics that emerged in the interviewees' discourse. We then take each point separately. Examples are numbered and the source of the data is provided at the end of the utterance, using the following notation: code of the religions group (see Tables 1 and 2) and either the number assigned to the interview or FG for focus groups.

7.1. Broad picture: what did the interviewees talk about?

Table 3 displays the 10 most salient keywords within each set of data, grouped by meaning/function and ordered by LR. As shown, religion-related words were frequent in the discourse. This is not surprising as the interviewees were selected for their religious identity and the main purpose of these interviews was to discuss the role of religion in their integration into the British society. The geographical references in the *background* and *host country* categories reflect the interviewees' mentions of their ancestral lands and the host country. The asterisk indicates that *born* fits in more than one category; some were born outside Britain, others were Britain-born. *Miscellaneous* gathers words that did not fall within any other category. As we shall see shortly, they relate to topics that the analysis of semantic domains revealed to be prominent in the discourse.

Category	Christians (CR)			Hindus (HD)			Muslims (ML)			Super Group (SG)		
	Keyword	Freq.	LR	Keyword	Freq.	LR	Keyword	Freq.	LR	Keyword	Freq.	LR
Religion	faith	114	5.3	temple	156	11.6	mosque	99	10.5	Islam	44	10.0
	Christian	56	5.1	Hindu	154	11.6	Islamic	86	10.3	Muslims	39	9.8
	church	117	4.7	Hinduism	37	9.5	Islam	164	10.3	Muslim	82	8.9
	God	105	4.3	beliefs	35	8.4	Muslims	152	10.1	religion	45	7.2
				religion	244	8.4	Allah	57	9.7			
				religions	39	7.6	Muslim	362	9.4			
				prayers	31	7.3						
Background	Caribbean	36	8.5	India	270	7.0	Gujarati	62	9.9	Pakistan	34	10.6
	born*	39	3.8				Bengali	43	9.3	Pakistani	30	9.4
										Asian	63	6.9
										India	45	5.6
Host country	born*	30	3.8	UK	41	9.7	Lancaster	115	9.7			
				Preston	30	9.2	Preston	36	9.1			
Miscellaneous	kind of	71	5.2							culture	33	6.4
	education	44	4.6							community	73	5.3
	parents	47	3.7									
	life	98	3.4									

Table 3: *Most salient words/phrases in the discourse of each religious group, where* Freq. *indicates raw frequencies*

Table 4 displays the five semantic domains most salient within each religious group, with their LR and the five most frequent words within each domain. Note that only two domains appeared in all four groups: *religion and the supernatural* and *kin*. However, most domains that did not figure in this table were just below the thresholds, including those that appeared in one group only. For example, the semantic domain of *hindering* was salient in CR but emerged in the 6th and 10th positions in the HD and ML rankings respectively. In the SG, it did not reach the minimum frequency threshold (50 instances). This means that Christians mentioned the obstacles that they faced more frequently than the other groups did. This was also the case of mentions of successful achievements (cf. *success* in CR) and of daily routine (cf. *frequency* in ML).

Christians (CR)	Hindus (HD)	Muslims (ML)	Super Group (SG)
Religion and the supernatural (LR = 3.9): faith, church, God, lord, religion	Religion and the supernatural (LR = 3.7): religion, Hindu, faith, God, religious	Religion and the supernatural (LR = 3.7): Muslim, religion, Islam, Muslims, mosque	Religion and the supernatural (LR = 3.4): Muslim, religion, Islam, Muslims, faith
Education in general (LR = 2.3): school, education, training, schools, teacher	Personal relationships – general (LR = 2.0): friends, meet, met, friend, get on	Education in general (LR = 2.2): school, college, education, university, schools	Geographical names (LR = 2.0): Asian, British, India, Pakistan, Pakistani
Hindering (LR = 2.0): obstacles, hindrance, obstacle	Education in general (LR = 2.0): school, education, university, schools, study	Kin (LR = 1.8): family, parents, dad, mum, children	Personal relationships – general (LR = 1.8): friends, friend, met, meet, get on
Kin (LR = 1.5): family, parents, kids, children, father	Geographical names (LR = 1.9): India, Indian, British, temple, Asian	Frequency (LR = 1.8): sometimes	Kin (LR = 1.8): family, parents, kids, father, grandfather
Success (LR = 1.5): success, successful	Kin (LR = 1.8): family, children, parents, father, husband	Personal relationships – general (LR = 1.8): friends, meet, friend, met, get on	People (LR = 1.7): people, person

Table 4: *The five most salient semantic domains by religious group, with the resulting LR and the most frequent words within each domain*

The analysis of semantic domain corroborates and supplements the tendencies revealed by the keywords. The high prominence of the domain of *religion and the*

supernatural is a clear indication that religion was a salient topic across all groups. The combination of the domains of *kin* and *geographical names* suggests frequent references to their ancestral lands and family roots. The host culture is also evident, as indicated by the word *British* within *geographical names*. Together with the words *friends* and *people* (the most frequent within the domain of *personal relationships and people*), the word *British* points towards frequent references of the British wider community. Lastly, *education* appeared as a salient topic because the interviewees were asked about their educational background and encouraged to talk about their experience through the British education system and their families' perspectives on education.

Tables 5 and 6 indicate differences across groups. They display the keywords and key semantic groups of each group in relation to a combination of the others.

Category	Christians (CR)			Hindus (HD)			Muslims (ML)			Super Group (SG)		
	Keyword	Freq.	LR	Keyword	Freq.	LR	Keyword	Freq.	LR	Keyword	Freq.	LR
Religion	church	117	4.3	temple	156	6.6	Islamic	86	6.6			
	Christian	56	2.8	Hinduism	37	4.5	hijab	34	6.3			
	word	32	2.8	Hindu	154	3.3	Quran	31	6.1			
							prophet	30	6.1			
							Allah	57	6.0			
							mosque	99	5.2			
							Islam	164	4.4			
Background	Caribbean	36	5.8	Indian	158	2.0	Bengali	43	6.6	Pakistan	34	2.6
										Pakistani	30	2.0
Host country							Lancaster	115	8.0			
Miscellaneous	black	57	2.8	hospital	38	2.7	wearing	37	6.4	and so on	31	4.0
				doctor	40	2.6				woman	36	2.4
				initially	35	2.4				interesting	33	2.3
				practice	32	2.3						
				food	36	2.1						

Table 5: *Keywords specific to a group;* Freq. *indicates raw frequencies*

Christians (CR)	Hindus (HD)	Muslims (ML)	Super Group (SG)
	Medicine and medical treatment (LR = 2.0): doctor, hospital, medical, medicine, surgery	Clothes and personal belongings (LR = 2.3): wear, wearing, hijab, scarf, clothes	
		People: Female (LR = 1.6): women, girls, girl, woman, ladies	

Table 6: Semantic domains specific to a given religious group, with the resulting Log-Ratio (LR) and the five most frequent words within each domain

Note that Table 3 and Table 5 share various keywords within the *religion, background* and *host country* categories. In fact, these results reinforce one another. Irrespective of how the data was approached, the keywords highlighted the self-reflexive nature of the narrative. Interviewees reflected upon their own individual religion. Christians talked about the *Church of England* and adherence to *the word of God* while Hindus mentioned *Hinduism* and their praying, and Muslims referred to the *Quran*, the *prophet* or *Allah*. Comprising mostly Muslims, the Super-Group frequently used the words *Islam, Muslim*, and *Muslims* (Table 3).

A similar trend was seen in relation to family background. Christians highlighted their Caribbean roots, Hindus made reference to India, and Muslims and the Super-Group mentioned the Indian subcontinent. There were also those from Pakistan, especially from the Super-Group; and a few from Bangladesh or surrounding countries. In some cases, the interviewees mentioned the specific regions where they came from (cf. Gujarati and Bengali, in Table 3).

The results also unveiled distinctive features in the discourse of each group. Christians were the only ones that mentioned their skin colour or race, as indicated by the word black (Table 5). The health-related words among Hindus (Tables 5 and 6) relate to the interviewees' professional lives; there were eight medical doctors and a nurse. Half of the occurrences of the word *food* (HD, Table 5) came from one single interview with someone that worked in the food sector. The words *wearing* in ML and *woman* in SG (Table 5), and the semantic domains *clothes and personal belongings* and *people: female* in ML (Table 6), both relate to the Muslim-female practice of wearing a headscarf (or a veil). They were used by female speakers to explain why they wore the hijab or to recount their experiences in breaking stereotypes.

Overall, this combination of approaches provided evidence to support the claim that the interviewees' discourse revolved around the same topics, irrespective of their religious faith or background. This was expected given that the interviews were designed to follow the same guidelines and achieve similar goals. At the same time, the overlaps suggest that different groups articulated similar ideas when talking about

their religion and interaction with the host country. We now turn to the most dominant aspects of the discourse: (i) religion; (ii) family origins and roots; (iii) the wider community; (iv) identity; and (v) education.

7.2. Religion

When referring to their religion, Christians showed a clear preference for the word *faith*, Hindus opted for either *religious beliefs* or *religion*, Muslims and the Super-Group talked about *Islam* or *religion* (Tables 3 and 4). Despite their different linguistic choices, the discourses were similar. The interviewees talked about their own religion; references to religions other than that of the speaker were scarce. This was seen through the frequent appearance of the pronouns *my*, *our*, and *your* (used impersonally to refer to one's religion in general) around the words *faith* in CR and *religion* in the other groups.

Most interviewees regarded their religion as a crucial aspect of their lives (Excerpts 1-2). They expressed how they related to their religious beliefs and negotiated its principles in their everyday life and interaction with other people. Their faith guided their behaviour and provided support and strength, especially when going through hard times.

1   … my religious belief means the world to me, it's my faith in God, it's my faith in God that has allowed me to succeed … (CR12).
2   My life is my religion. It's everything. I would say it's really important because we've been told that if Islam is a part of life … (ML-GJ03)

Some explained their understanding and interpretation of their religious principles and/or how they complied with those principles. This last aspect was especially evident among Hindus, and HD-SI in particular, who viewed Hinduism as a way of life, with no fixed beliefs or strict rules of conduct and hence flexible and liberal in its basic principles. Many interviewees, especially in the HD-SI and the SG, stated that religion cannot be seen in isolation; it is intertwined with one's identity, culture, and family values (Excerpt 3).

3   No everything comes into it you know, your success in life, philosophy, religion. Because I don't think anything in the world works in isolation really… (SG04).

Lastly, all groups but Christians mentioned the issue of discrimination. Hindus and the Super-Group tended to talk about it in general terms, expressing disapproval but making no reference to a specific religion. Muslims, especially ML-GJ, specifically mentioned prejudice against the Islamic religion and culture. Some referred to the ways in which the British media has negatively portrayed Muslims and by thus influencing public opinion and creating disharmony among those who do not share similar religious beliefs.

7.3. Family origins and roots

The interviewees' bond to their family roots was clearly seen through mentions of geographical place-names. Christians frequently emphasized that they were African *Caribbean* (Table 3). Hindus and Muslims tended to establish a direct link between religion and their country of origin by referring to themselves as, for example, *Indian Hindus* or *Asian Muslims*. Such association between origins and religion was not found in the Super-Group. However, all three groups (HD, ML and SG) expressed strong connection with the Indian or Pakistani cultures (Tables 3 and 4).

The regular appearance of the word *community* around the words *Hindu*, *Muslim, Indian,* or *Asian* in the HD, ML and SG groups revealed the interviewees' strong sense of community, especially towards those who shared their identity, culture, and religion. This aspect was especially evident among HD-GJ who frequently mentioned their *community centre*, which facilitates and enhances connections between members of the local Hindu community. Among Christians, this sense of community came from the word *church*, which occurred with a much higher relative frequency than its equivalents *temple* and *mosque* in the other groups. Many expressed engagement with their (*local*) *church community*, which they usually met on Sunday services.

Acknowledgement of their family roots also came from the word *family*. In addition to mentioning their family's ancestral lands, all groups reported preserving their family traditions, values, religious beliefs and practices. Some referred to their family as a source of support and help while others recounted personal experiences, explaining why and how the family had immigrated to Britain.

7.4. The wider community

While showing active engagement with their own community, many interviewees stressed that it is essential to integrate and contribute to the *host community* or the *British community* (Excerpt 4). This feature was less evident among HD-GJ and CR. The former hardly mentioned any interaction with the wider community while the latter only occasionally referred to those living nearby, who may not necessarily be Christians.

> 4   So the actions I do and my beliefs along with the British community, I wouldn't try to alienate myself away from the community. I would like to be part of the community and integrate well into the community (HD-SI03).

Through the word *friends*, we found that interviewees from all groups mingled and interacted with people from various backgrounds and religions, an important element in their integration into the British society. But interviewees also acknowledged friendship with those who shared their religious beliefs and family roots. This aspect was especially evident among Muslims; over half of them stressed their closest friends were Asian Muslims.

7.5. Identity

Most interviewees regarded themselves as British, irrespective of religion or birth place. The word *British* was highly salient across all groups and frequently preceded by the first person pronoun (*I*) and either the verb to be (*am*) or verbs such as *feel*, *say*, or *consider/see/class/categorise myself as*. Only three out of the total number of interviewees stated they did not feel British: one in each group, except in the SG.

What is more, interviewees did not see themselves as solely British but, rather, their identity lay at the intersection of two or more identities (Excerpts 5-8). Christians tended to refer to themselves as *British Caribbean*; some as *Black British*, thus projecting their race into their identity. Hindus considered themselves *British Hindus* or *British Indians* or *Asians*. Muslims and many within the Super-Group regarded themselves as *British Muslim*. Some were more specific, calling themselves *British Asian*, *British Indian* or *Bengali Muslim British* or *British Pakistani*.

> 5   Essentially I call myself African Caribbean but I'm a British citizen (BC05).
> 6   So I am a Hindu and I'm a British Hindu. I'm comfortable to be a British Hindu, I 'm
>   comfortable to be Indian Hindu (HD-SI09).
> 7   Gujurati, Indian Muslim. British, Indian Muslim. I've got many identities (ML-GJ04).
> 8   I would say I'm a British Pakistani Muslim, that how I describe myself (SG11).

The discourse provides evidence of the interplay between their religious faith, ancestral origins, and the host culture. Essentially, it shows that the interviewees were aware of their nested position in negotiating and bridging across different contexts, cultures and communities. Nearly all interviewees highlighted that the British culture was blended with their ancestral roots. Many offered their reasons for feeling British. Some were born and raised in Britain, others were born somewhere else but either brought up in Britain or had lived in Britain for a long time. Some established a hierarchical order for the components of their identity, with either their birth place or religion coming first, and Britain second. For others, their British identity was associated with citizenship and having the British passport.

In many cases, the interviewees broadened the discussion out by reflecting upon what the British identity entails (Excerpt 9). For some, it relates to a feeling of belonging to culture and society whose values they recognise and share. For others, the British identity congregates and embraces various identities; a pluralistic society that accommodates a wide range of cultures, religions and practices, thus allowing them to feel accepted and integrated.

> 9   I think citizenship is about a commitment to the space, the culture and institutions that
>   you share and so I consider myself as a British citizen, identifying with the
>   multiculturalism that is modern Britain you know (CR-01).

Few interviewees reported having experienced some kind of discrimination by having their British identity challenged, be it for their skin colour, religion or ancestral culture. This feeling of discrimination pushed some (approximately 20% to 30% of cases within each group) closer to their homelands or ancestral roots despite their British citizenship.

7.6. Education

The frequency of the word *education* – 4.8 occurrences per 10,000 words in the ML-GJ and at least 6.1 in all other groups – mirrored the interviewees' level of qualification. As mentioned earlier (Section 2), most interviewees were highly educated; with the exception of those within ML-GJ.

Interviewees from all but the ML-GJ group mentioned their qualifications and educational background. They viewed education as crucial for taking one forward and succeeding in life, especially for immigrants. For some, education was a key reason for themselves or their parents to have immigrated to Britain. There were also accounts of personal experiences, including the impact of education on their personal lives, how and where they were educated, and culture clashes or discrimination they faced.

8.   What we learned

This is an unusual dual study, approaching a research question with two contrasting methodologies. As an investigation of the views of religious and ethnic minorities, the findings are consistent with each other as well as with the current, general view of minorities in Britain. The results corroborate the argument that religion provides immigrants and second and third generations not only with spiritual guidance and support, but also with a sense of belonging and identity. The interplay between religion and culture was evident; religion was viewed as a way of life and a key element in determining how one chooses to lead life, relate to others, and preserve roots, traditions and values. Immigrants use their places of worship to practice their religion and as repositories for experiencing and preserving their traditions and cultural values by mingling and interacting with those from similar background and faith (Chivallon, 2001; Geaves, 2015; Maliepaard and Schacht, 2018; Sharma, 2012; Werbner, 2002, pp. 255-256).

This study also extends the discussion around immigrants' identity by providing evidence of the interviewees' blend of identities, irrespective of their religious faith, background or generation. Scholars have claimed that the level of religiosity declines among the second and third generation immigrants (Lewis and Kashyap, 2013) and so does their bonding with their family roots as their priorities and interests merge traits of the host and the ancestral cultures (Sharma, 2012). In relation to British Muslims specifically, the literature argues that, after the 9/11 and the 7/7 events, British-born Muslims began to view themselves as a mix of the two cultures. While defending Islam and Muslim values, they associated themselves with the British culture and society (Geaves, 2015; Lewis, 2007, p.149; Sharma, 2012).

The literature says little about levels of education among religious immigrants. One exception is Sharma (2012), who states that the level of education among the Indian diaspora is higher than other immigrant communities in Britain. Indians occupy permanent positions in reputed universities, own big- and medium-sized enterprises or are employed in various service sectors. The present analysis contributes by highlighting that immigrants view education as a key element for their successful integration into the British society.

As for the comparative question of what qualitative analysis and corpus techniques can say to one another, our overall conclusion is that the analysis was mutually informative, offering both fields valuable lessons.

The use of corpus techniques did yield an empirical backing and fresh perspectives to the qualitative study. The qualitative, hermeneutical interpretive method (consciously) predisposed the original researchers to follow the interviewee's lead, and unpack their world of meaning through their own idiosyncratic speech. Corpus techniques on the other hand allowed observations of which the participants were unaware; with an empirical impulse not always given by qualitative analysis. The corpus analysis picked up linguistic patterns that were not identified through the readings of single-texts. For example, the pronouns *my*, *our*, and *your* around religion-related words indicated that the interviewees reflected upon their own individual faith. Another example is the word *family* which uncovered mentions of family origins and roots, as well as traditions, culture, values, and religious beliefs.

Corpus techniques also provided compelling evidence to support the results of the initial qualitative analysis regarding common discourse patterns across all groups, regardless of background. The corpus analysis reiterated the initial conclusion that religious tradition or background played a key role in motivating and shaping the interviewees' behaviour, practices and relationships, supporting and helping them overcome difficulties and achieve success. Interviewees had a strong sense of community in relation to those with similar religious faith, background and identity as well as the British community at large. Both analyses demonstrated that interviewees accommodated their religion, ancestral lands and/or ethnicity into the British identity; only a few interviewees felt discriminated and not accepted as British, most often on racial or religious grounds.

However, whereas the initial qualitative analysis noted that many interviewees felt that they contributed to their local and/or the wider community in one way or another, this aspect of the discourse was not picked up by the corpus analysis. Nor did the corpus analysis surface the nuanced, ambivalent discussions about whether one should engage with 'the system' or 'the establishment', be a change-maker, overcome disillusionment and persist. This was perhaps because these aspects of the discourse were expressed in more complex linguistic ways which would not show up where the methodology is predicated upon repeated patterns of language (Baker, 2006, pp.13–14). It may as well relate to the establishment of cut-off points, which means some findings go unreported (Egbert and Baker, 2016).

The investigation also throws up some challenges for corpus linguistics. One challenge relates to the risk of cherry picking, which is a frequent issue when the researchers have pre-existing views on the data (Ancarno, 2018). The limited size of the corpora and distinctive nature of the data also raised many methodological questions. The analysis thus entailed a careful evaluation of the corpus techniques to ensure that they were appropriately applied in order to meet the intended purposes of the present study.

One important decision was to approach the data using various techniques. As Egbert and Baker (2016) observe, methodological triangulation allows the researcher to provide a more complete picture of the discourse as well as to validate the analysis

by cross-checking the results. In the specific case of this study, the combination of keyword analysis with key semantic domains has proved especially useful. As one procedure reinforced, complemented and supplemented the other, this combined approach boosted our level of confidence in deciding on which words had the potential to yield the most relevant results.

Another critical issue was to establish the thresholds for generating keywords and semantic domains (see Section 6). A particular problem in small corpora is that frequencies may be too small for results to be reliable, especially when it comes to statistical significance (Baker *et al.* 2008). Here, there was the additional issue that the resulting LR scores varied considerably across the subcorpora as well as between the calculations of keywords and semantic domains within the same subcorpus. This made it difficult to find a cut-off point that would be suitable for all calculations. Our decision to concentrate on a set number of keywords and semantic domains was thus informed by careful evaluation of different thresholds, with special attention to what would be overlooked in each case.

The reduced size of the corpus also guided our decision not to undertake collocation analysis. In a pilot collocational analysis of the most salient words, most collocates were function words (mainly pronouns, but also prepositions and determiners), linking and/or auxiliary verbs (especially *to be*, *have* and *will*). This is not to say that those collocates had no potential to yield interesting results but, rather, that they would require close reading of the extended context to be adequately interpreted. The analysis thus gained in quality through close reading of all instances of the selected words. We also examined the distribution of instances across texts to avoid highlighting patterns restricted to one (or few) individuals.

This study thus offers a significant contribution to the field of corpus linguistics from a methodological standpoint. It also adds to the scarce literature on corpus studies based on small corpora and those drawn on data collected for purposes other than linguistic research, and transcribed interviews in particular.

9. Conclusion

All in all, this study serves as a robust indication of the strength of corpus analysis in the limited case. By adjusting the techniques to adequately analyse the available material, the corpus analysis reiterated and complemented the results from the initial qualitative analysis of the data. At the same time, it was essential to combine knowledge from both fields in a way that the two approaches were mutually informative and the results appropriately interpreted in order to integrate corpus linguistics methods into a prototypical humanities study.

To conclude, it is reasonable to assume that with a larger set of data that is more difficult to cover through traditional analysis, corpus linguistics can do more than just conform to traditional techniques. A corpus analysis has the potential of yielding additional insights. At the very least, the method is especially useful to obtain an overall picture of the data and by thus putting the researcher in a stronger position to decide which parts of the data are worth closer examination. The present discussion is

thus relevant to any study that seeks to use corpus methods to explore data collected through standard social science methods, and interviews in particular.

Funding

Competing interests

The authors have no competing interests to declare.

References

Ancarno, C. (2018). Interdisciplinary approaches in corpus linguistics and CADS. In C. Taylor & A. Marchi. (Eds.), *Corpus Approaches to Discourse: a critical review* (pp. 130-156). London: Routledge.

Baker, P. (2006). *Using Corpora in Discourse Analysis*. London: Continuum.

Baker, P., Gabrielatos, C., Khosravinik, M., Krzyzanowski, M., McEnery, A. M. & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse and Society*, *19*(3), 273–306. doi:10.1177/0957926508088962

Baker, P., Gabrielatos, C. & McEnery T. (2013). *Discourse analysis and media attitudes: the representation of Islam in the British press*. Cambridge: Cambridge University Press.

Bauer, E. (2018). Racialized citizenship, respectability and mothering among Caribbean mothers in Britain. *Ethnic and Racial Studies*, *41*(1), 151–169. doi:10.1080/01419870.2017.1317826

Blinder, S. & Allen, W. L. (2016). Constructing immigrants: Portrayals of migrant groups in British national newspapers, 2010–2012. *International Migration Review*, *50*(1), 3–40, doi:10.1111/imre.12206.

Brezina, V., McEnery, T. & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, *20*(2), 139–173. doi:10.1075/ijcl.20.2.01bre

Brookes, G. & Baker, P. (2017). What does patient feedback reveal about the NHS? A mixed methods study of comments posted to the NHS Choices online service. *BMJ*

Open, 7:e013821. doi:10.1136/bmjopen-2016-013821 PMid:28450463 PMCid:PMC5566587

Calder, G., Cole, P. & Seglow, J. (2010). *Citizenship acquisition and national belonging: migration, membership and the liberal democratic state*. Basingstoke: Palgrave.

Capdevila, R. & Callaghan, J. E. M. (2008). 'It's not Racist. It's Common Sense'. A Critical Analysis of Political Discourse around Asylum and Immigration in the UK. *Journal of Community & Applied Social Psychology*, *18*, 1–16. doi: 10.1002/casp.904

Chivallon, C. (2001). Religion as space for the expression of Caribbean identity in the United Kingdom. *Environment and Planning D: Society and Space*, *19*(4), 461–483. https://doi.org/10.1068/d2341

Egbert, J. & Baker, P. (2016). Research Synthesis. In P. Baker and J. Egbert (Eds), *Triangulating Methodological Approaches in Corpus-Linguistic Research* (pp. 183-208). London: Routledge.

Ford, R. (2011). Acceptable and Unacceptable Immigrants: How Opposition to Immigration in Britain is Affected by Migrants' Region of Origin. *Journal of Ethnic and Migration Studies*, *37*(7), 1017–1037. doi:10.1080/1369183X.2011.572423

Gilquin, G., De Cock, S. & Granger, S. (2010). *The Louvain International Database of Spoken English Interlanguage. Handbook and CD-ROM.* Louvain-la-Neuve: Presses universitaires de Louvain.

Geaves, R. (2015). Review of the Contemporary Literature on Islam and Muslims in the UK through the Lens of Immigration Issues, Civic Participation and International Constraints. *Fieldwork in Religion*, *10*(1), 10–42. doi:10.1558/firn.v10i1.28330

Hardie, A. (*forthcoming*). A dual sort-and-filter strategy for statistical analysis of collocation, keywords, and lockwords. An informal introduction can be found at http://cass.lancs.ac.uk/?p=1133; accessed 2020-04-01.

Kaufmann, E. & Harris, G. (2015). "White Flight" or Positive Contact? Local Diversity and Attitudes to Immigration in Britain. *Comparative Political Studies*, *48*(12), 1563–1590. doi:10.1177/0010414015581684

Lewis, P. (2007). *Young, British and Muslim*. London: Continuum.

Lewis, V. A. & Kashyap, R. (2013). Piety in a Secular Society: Migration, Religiosity, and Islam in Britain. *International Migration*, *51*(3), 57–66. doi:10.1111/imig.12095

McEnery, T. & Baker, H. (2016). *Corpus Linguistics and 17ᵗʰ Century Prostitution.* London: Bloomsbury Academic.

McEnery, T., Love, R. & Brezina, V. (2017). Introduction: compiling and analysing the Spoken British National Corpus 2014. *International Journal of Corpus Linguistics*, *22*(3), 311–318. doi:10.1075/ijcl.22.3.01mce

Maliepaard, M. & Schacht, D. D. (2018). The relation between religiosity and Muslims' social integration: a two-wave study of recent immigrants in three European countries. *Ethnic and Racial Studies*, 41(5), 860–881. doi:10.1080/01419870.2017.1397280

Nolte, M. I., Ancarno, C. & Jones, R. (2018). Inter-religious relations in Yorubaland, Nigeria: corpus methods and anthropological data. *Corpora*, *13*(1), 27–64. doi:10.3366/cor.2018.0135

Partington A., Duguid, A., & Taylor, C. (2013). *Patterns and meanings in discourse theory and practice in corpus-assisted discourse studies (CADS)*. Amsterdam: John Benjamins Publishing Company.

Pérez-Paredes P., Jiménez P. A. & Hernández P. S. (2017). Constructing immigrants in UK legislation and Administration informative texts: A corpus-driven study (2007–2011). *Discourse & Society*, *28*(1), 81–103. doi:10.1177/0957926516676700

Potts A. & Kjaer, A. L. (2016). Constructing achievement in the International Criminal Tribunal for the Former Yugoslavia (ICTY): A corpus-based critical discourse analysis. *International Journal for the Semiotics of Law*, *29*(3), 525–555. doi:10.1007/s11196-015-9440-y

Prentice S., Rayson, P., & Taylor, P. J. (2012). The language of Islamic extremism: Towards an automated identification of beliefs, motivations and justifications. *International Journal of Corpus Linguistics*, *17*(2), 259 –286. doi:10.1075/ijcl.17.2.05pre

Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics 13*(4), 519–549. doi:10.1075/ijcl.13.4.06ray

Rayson, P. (2009). *Wmatrix: a web-based corpus processing environment*. Lancaster: Computing Department, Lancaster University.

Sealey, A. (2009). Probabilities and surprises: a realist approach to identifying linguistic and social patterns, with reference to an oral history corpus. *Applied Linguistics*, *31*(2), 215–235. doi:10.1093/applin/amp023

Sealey, A. (2012). 'I just couldn't do it': representations of constraint in an oral history corpus. *Critical Discourse Studies*, *9*(3), 195–210. doi:10.1080/17405904.2012.688295

Sharma, S. (2012). Immigrants in Britain: A Study of the Indian Diaspora. *Diaspora Studies*, *5*(1), 14–43. doi:10.1080/09739572.2013.764123

Small, S. & Solomos, J. (2006). Race, Immigration and Politics in Britain: Changing Policy Agendas and Conceptual Paradigms 1940s–2000s. *International Journal of Comparative Sociology*, *47*(3-4), 235–57. doi:10.1177/0020715206065781

Smith, J., Flowers, P., & Larkin, M. (2009). *Interpretative Phenomenological Analysis: Theory, Method and Research.* London: Sage Publications.

Van Dijk, T. (1991). *Racism and the Press*. London: Routledge.

Webber, F. (2014). Extending immigration policing and exclusion in the UK. *Race & Class 55*(3), 86–92. doi:10.1177/0306396813509199

Werbner, P. (2002). *Imagined Diasporas among Manchester Muslims*. Oxford: James Currey.