From *gay language* to *normative discourse*

A diachronic corpus analysis of Lavender Linguistics conference abstracts 1994–2012

Paul Baker
Lancaster University, United Kingdom

A corpus of abstracts from the *Lavender Languages and Linguistics Conference* was subjected to a diachronic keywords analysis in order to identify concepts which had either stayed in constant focus or became more or less popular over time.[1] Patterns of change in the abstracts corpus were compared against the *Corpus of Contemporary American English* (COCA) in order to identify the extent that linguistic practices around language and sexuality were reflected in wider society. The analysis found that conference presenters had gradually begun to frame their analyses around queer theory and were using fewer sexual identity labels which were separating, collectivising and hierarchical in favour of more equalising and differentiating terminology. A number of differences between conference-goers' language use and the language of general American English were identified and the paper ends with a critical discussion of the method used and the potential consequences of some of the findings.

Keywords: abstracts, corpus, language, sexuality, diachronic

1. Introduction

At the time of writing, since 1993, the *Lavender Languages and Linguistics* (LavLangs) conference has been an (almost always) annual event, taking place at American University, Washington DC and organised by William (Bill) Leap, a Professor of Anthropology from the same institution. The conference has overseen the birth of the field of Language and Sexuality as well as providing an international forum for steering its development and enabling cross-pollination of ideas from a diverse range of fields which include Linguistics, Languages, Literature, Women's Studies, Lesbian and Gay Studies, Queer Studies, Anthropology, Sociology, Media

Journal of Language and Sexuality 2:2 (2013), **179–205**. DOI 10.1075/jls.2.2.01bak
ISSN 2211–3770 / E-ISSN 2211–3789

and Film. At the conference website, under the heading *Conference Focus*, the aims of the conference are described as follows:

> Stated broadly, the Lavender Languages and Linguistics Conference examines language use in lesbian, gay, bisexual, transgender, and/or queer life. Linguistic inquiry is broadly defined here, to include studies of: pronunciation, vocabulary and meaning, conversational structures and styles, life stories and other narratives, fiction, and poetry, the "language" of scientific and historic documents and print media, meanings encoded in spatial practices, sign language, non-verbal communication, and communication through photography, cinema and other visual arts. (Lavender Languages 2013)

As the conference approached its 20th anniversary in 2013, I felt that it was pertinent to consider how the field of Language and Sexuality has developed, particularly as the last two decades have seen unprecedented changes in the social and legal status of LGBT people across the world. While such changes have generally been hailed as positive in relatively rich western countries, the situation is more complex for other parts of the world, and in some cases has measurably worsened. One research question that this paper wishes to address is: to what extent is the language use of conference-goers reflected in the language of general society over time? However, I also wish to use this paper as a way of focussing on the conference itself, particularly taking into account the fact that the conference has sometimes played host to lively debates about the way that research in the field of Language and Sexuality should be carried out, which concepts should be given precedence and indeed, whether certain concepts actually exist at all. If we were to examine the language that has been produced by conference-goers over time, what would it tell us about changing themes and foci? In order to answer these questions, a corpus linguistics approach has been undertaken, allowing us to examine how language use has altered or stayed the same since the earliest years of the conference. This method was chosen as first, corpus linguistics is ideally placed for identifying linguistic patterns and trends across large amounts of data that would be difficult to spot by hand and eye methods, and second, corpus linguistics reduces researcher bias by presenting analysts with linguistic items that are statistically salient or frequent — items which may run counter to a researcher's own hypotheses or hunches. While corpus techniques do not completely remove researcher bias (the items still have to be qualitatively interpreted and explained, and in some cases certain items may receive more analytical focus than others), they do reduce bias considerably, and as someone who wishes to take a reasonably objective overview of how the field has developed, I feel that this approach works well.

In this paper, after discussing some of the research which has influenced the approach that was taken here, I move on to describe how the corpus of conference

abstracts was collected, complied and interrogated. This is followed by the analysis of the corpus, centred around the research questions outlined in the above paragraphs. The paper concludes with some critical notes about the method used and its limitations, followed by a discussion of emerging themes from the conference.

2. Literature review

It is not the intention of this literature review to summarise what other people have said about the development of the field of Language and Sexuality (although key texts include Cameron 2005, Cameron & Kulick 2003, Campbell-Kibler, Podesva, Roberts & Wong 2002, and Motschenbacher 2012). Instead, I wish to discuss an area which is perhaps less familiar to researchers in this field, that of Corpus Linguistics and how it can be used to investigate linguistic and discursive change over time.

Research into diachronic change in Corpus Linguistics initially tended to focus on issues pertaining to grammar. For example, Smith (2002) examined change in the use of the progressive, comparing two corpora each consisting of one million words of written published British English: LOB (containing texts from 1961) and FLOB (containing texts equivalent to FLOB from 1992). Smith found that generally, there were more progressives in the FLOB corpus, which was particularly due to a rise in the use of the present progressive. A similar study by Leech (2003) used the same corpora (and their American equivalents) in order to examine the development of modal verbs. Leech argues that generally, modals appear to be declining, with the pattern being more advanced for American English. However, Millar (2009) raises a note of caution, indicating that it is dangerous to make strong interpretations when there are only two sampling points. The corpora tell us nothing about what happened before, after or in between these points, and Millar's examination of modals in the more fine-grained TIME corpus (all of the yearly output of *Time* magazine since 1923) actually found evidence that modals had increased. A response by Leech (2011) to Millar's paper used five sampling points rather than two, still finding evidence that modals were decreasing. Perhaps Millar's conflicting findings are due to the TIME corpus being genre-specific rather than fully representative of written language. However, this series of studies highlights the importance of aiming for a high granularity (the number of sampling points divided by the span of years across the corpora being considered), as well as taking care not to generalise findings beyond the genre(s) of the corpora being examined.

Another aspect of diachronic corpus research has investigated cultural change. Taken together, a series of studies (Leech & Fallon 1992, Oakes 2003, Potts & Baker 2012) have compared pairs of American and British corpora from 1961, the early 1990s and 2006 respectively, using statistical tests in order to identify

differences in lexical frequency between the two language varieties. Potts and Baker (2012) provide an overview of differences across the three time periods, noting that some differences appear to have remained stable across the two countries (British English always contains more words relating to time and modality, while American English always has more words to do with the military and computing). However, other patterns show changes — with American English having more references to excessive drinking in the 1960s, while in the 2000s, it is British English which appears to be more focussed on this concept.

Looking closely at gender (and particularly male bias), Holmes and Sigley (2001) and Sigley and Holmes (2002) used the 1960s and 1990s corpora discussed above to examine frequencies and contexts of nouns which identified people by their sex (*man*, *woman*, *boy*, *girl*, *lady* and their plural forms), including nouns which contained gender-marking suffixes like *–ess* and *–ette*. They found reductions in frequencies of some sexist terms in the later period such as the pseudo-polite *lady*, generic uses of *man*, and the *-ess* and *–ette* suffixes. Baker (2010) updated this research to also consider equivalent corpora from the 1930s and 2006, finding evidence that over time there only seems to have been a slight uptake of linguistic strategies to remove male bias from language such as inclusive pronouns, *Ms* or gender-neutral terms like *spokesperson*. Instead, an overall decrease in male pronouns was found, as well as a sharp decline in *Mr*. Baker suggests that one way that the gender inequalities surrounding the English term of address system (*Mrs/Miss* vs. *Mr*) could be resolved is that people will simply abandon the term of address system altogether, rather than taking on *Ms* as an equivalent to *Mr*.

A related perspective on diachronic change in corpora has been taken by the CADS (Corpus Assisted Discourse Studies) group, headed by Alan Partington. In 2010 a special issue of the journal *Corpora* was devoted to the field of Modern Diachronic Corpus-Assisted Discourse Studies. The papers compared corpora of British broadsheet newspapers from 1993 and 2005, looking at phenomena including informalisation of writing style (Duguid 2010), markers of evidentiality (Clark 2010), discourses surrounding morality (Marchi 2010) and constructions of science (Taylor 2010). An important technique involved the elicitation and analysis of keywords (see for example, Duguid 2010) where the frequencies of all the words in two corpora are compared together and either log-likelihood or chi-square tests are used in order to identify words where there is a statistically significant difference. Keywords can act as signposts, offering researchers a way to narrow their focus towards salient phenomena in their data that they perhaps would not have otherwise considered to be important. Popular corpus tools like WordSmith (Scott 2008) and AntConc (Anthony 2011) allow keywords to be derived automatically.

A pertinent study for this research is Johnson (2012), who examined one and a half million words of papers published in the *International Journal of Corpus*

Linguistics, from the period 2000–2011, splitting her data into three time periods (labelled early, mid and late). Johnson used the keywords technique to identify words which were statistically salient in each of the three periods, finding that earlier articles in the journal tended to show a focus on translation, terminology and collocation whereas as the field developed, more attention was paid to discourse, grammar and spoken language.

Finally, a study which used a similar method to that in this paper (although not based on academic data), is reported in Baker, Gabrielatos and McEnery's (2013) discussion of change over time across British newspaper articles on the topic of Muslim and Islam. Based on a technique used by Culpeper (2009) who compared the speech of characters in *Romeo and Juliet*, the newspaper corpus was divided into eleven periods (each consisting of a year from 1998 to 2009). The word lists (list of the frequencies of each word) for each individual year were compared against a wordlist consisting collectively of all of the remaining text in the corpus (e.g. 1998 was compared against 1999–2009). This technique produced a set of keywords for each time period, allowing keywords to be compared to see whether they had fallen in or out of fashion. For example, keywords relating to sexuality and gender were often found to occur during the period 1998–2000, although they were not present from 2001–2005 when instead there were keywords relating to terrorism, perhaps suggesting that one sort of news story had been replaced by another. By 2006–2009, however, the sexuality and gender keywords had reappeared, possibly indicating how a major event like 9/11 can help to 'derail' discussion of certain topics for a short period of time.

3. Method

Following Johnson (2012), compiling a corpus consisting of papers from this journal would be ideal in order to examine change in the field of Language and Sexuality. However, as this journal only began publishing in 2012, there would not be enough data to examine change over time or indeed to obtain a general impression of the field. Therefore, it was decided to focus on language use from the annual Lavender Languages and Linguistics conference, which has the distinction of dating back to 1993. It was not possible to obtain full copies of actual conference papers, but abstracts from some of the years were available for download in pdf form at the conference website. While the conference abstracts were typically quite short, consisting of only a few hundred words, they tended to provide succinct summaries of a paper's focus, research questions and theoretical and methodological frameworks used, so they could be classed as a 'rich' source of data.

After contacting the conference organiser, versions of most of the remaining abstracts were obtained. It was not possible to gather abstracts for 1993 or 2006, and the conference did not run in 2001 so not every year is represented. The set of abstracts that had been collected were 'cleaned' by removing unwanted text. This included names and affiliations of speakers as well as reference lists. In order to make the task of comparing different time periods easier, abstracts from adjacent years were joined together. This reduced the number of time periods to examine from 17 to a more manageable 8, putting more text in each time period and reducing the chance that a couple of articles appearing in the same year on a single topic would skew frequencies in their favour. The amount of data for 2003, 2004 and 2005 was relatively sparse, so these three years were joined together. In all, each of the 8 sections contained 15,000–23,000 words, while the total corpus size was 158,000 words.

As stated in the previous section, other researchers had used the keywords technique in order to compare multiple sets of corpora. Baker, Gabrielatos and McEnery (2013) had derived keywords for each year of a corpus of newspaper articles by using the 'remainder' of the corpus as the reference. This technique was initially used with the conference abstracts, although due to the relatively small amount of data being compared, it tended to produce a very small number of keywords for each period (between 7 and 22) using WordSmith's default settings ($p > 0.000001$). This is not necessarily problematic but the keywords produced in this way tended to focus on relatively rare concepts that were unique to a particular time period. For example, one keyword for the period 1998–1999 was *Cathal* (due to three papers in 1998 being about the poet Cathal O'Searcaigh). While interesting, this form of analysis tended not to produce more general words relating to over-arching themes which were better spread across a wider number of papers. For this reason, it was decided to adapt the method to instead compare each of the time periods against a much larger corpus, which would stand as reference for general English. I chose the AmE06 Corpus (Potts & Baker 2012), which was built using the same sampling framework as the LOB and FLOB corpora discussed earlier. The AmE06 corpus contains one million words of published written text mainly from 2006, in American English. As the LavLangs conference has always taken place in America, large numbers of attendees are American, and American English tends to be a popular form of language taught to overseas learners of English, it was felt that this corpus would act as an appropriate standard reference.

Using the default settings in WordSmith again, several hundred keywords were produced for each time period, and in order to focus the analysis on a manageable number of words, it was decided to only consider 100 keywords with the highest keyness values from each time period. These keywords were transferred to an Excel spreadsheet, allowing comparisons to be made across each time period.

Table 1. Sample keywords.

	1994–5	1996–7	1998–9	2000–2	2003–5	2007–8	2009–10	2011–12
analysis					✓	✓	✓	✓
coming	✓	✓						
discourse	✓	✓	✓	✓	✓	✓	✓	✓
ethnographic							✓	

Table 1 shows a small sample of the keywords, for illustrative purposes. A tick indicates that a word was key in that time period, when the abstracts for that time period were compared against the AmE06.

As a first step, the keywords were grouped into patterns based on when and how often they were key. For example, in Table 1 the word *analysis* was key in the four latter periods under examination, but not in the first four periods. Therefore, this word appears to have become more significant to abstract writers after 2003. The word *coming* (almost always used in the phrase *coming out*) was only key in the first two periods, and never key again, suggesting a word which has dropped out of fashion. *Discourse*, on the other hand was key in all eight time periods, indicating a concept that has held strong across the entire period that the conference has been running. Finally, *ethnographic* was only key in one time period. Although this is one of the later periods, the fact it only occurs once is perhaps less indicative that this is a word that is becoming more popular over time.

A set of criteria was devised in order to group keywords. Words that were key in all eight time periods were placed into one group as being 'always key'. Words that were key in at least three of the latter four periods and in only two or fewer of the first four periods were viewed as 'recently key'. Conversely, words that were key in at least three of the former four periods and in two or fewer of the latter four periods were viewed as 'previously key'. Finally, words which were key in only two periods (either in the first or second set of time periods) but nowhere else, were viewed as either 'recently key' or 'previously key', respectively. Using these criteria, Table 2 shows the different sets of words, which have been further categorised according to the concepts they referred to.

Table 2 shows concepts which have tended to remain a constant topic of focus, those which appear to be losing ground and those which have become more popular. However, the table does not explain why these words appear in certain positions, nor does it give any idea with regard to the contexts that they are used in abstracts. For this reason, the words in the table were subjected to concordance analyses (involving looking at all of the citations of each word across the whole set of abstracts, in context), in order to obtain a better idea about how and why they were used, and whether their meanings or associations had changed over time.

Table 2. Keyword classification.

Concept	Always key	Previously Key	Recently Key
Sexuality	gay, lesbian, queer, sexual, sexuality	homosexual, homosexuality, gays, bi, orientation	LGBT, homophobic, homophobia
Gender	male, gender	transsexual	transgender, trans, gendered
Analytical foci	discourse, language	linguistics, representations	discursive, analysis, presentation
Theoretical foci	identity, identities	culture, desire	normative
Grammatical words	of	which, within	in, non

In some cases, additional sources of information from outside the corpus were obtained, in order to shed further light on particular patterns.

In order to compare usage to the wider general population, another corpus was consulted — the COCA (*Corpus of Contemporary American English*), created by Mark Davies at Brigham Young University which can be accessed for free via an online interface (http://corpus.byu.edu/coca/). It consists of 450 million+ words of text from speech, fiction, magazines, newspaper and academic writing that were produced between 1990 and 2010. The COCA is therefore a good comparator to the corpus of abstracts as it covers a similar time period and is from the same variety of English. The following analysis sections are centred around four of the rows of Table 2: sexual identity words, gender and analytical and theoretical concepts. I did not focus on the grammatical words as they were generally more revealing of small stylistic differences which might be related to changes over time in general English (e.g. *within* being replaced by a shorter equivalent: *in*).

4. Analysis

4.1 Sexual identity words

Table 2 shows that some sexual identity words have altered in terms of popularity over time, while others have always remained key. Figure 1 shows the frequencies per 100 words of a number of these words (singular and plural forms were added together where applicable) over the different time periods of the corpus, including a couple of additional ones which were never key but for the purposes of making a fuller comparison are interesting to examine.

The figure shows that at the earliest period of the conference, *lesbian(s)* was the most common sexual identity word (of those in the graph), although this term has been on a reasonably steady decline since then (even though it has always remained

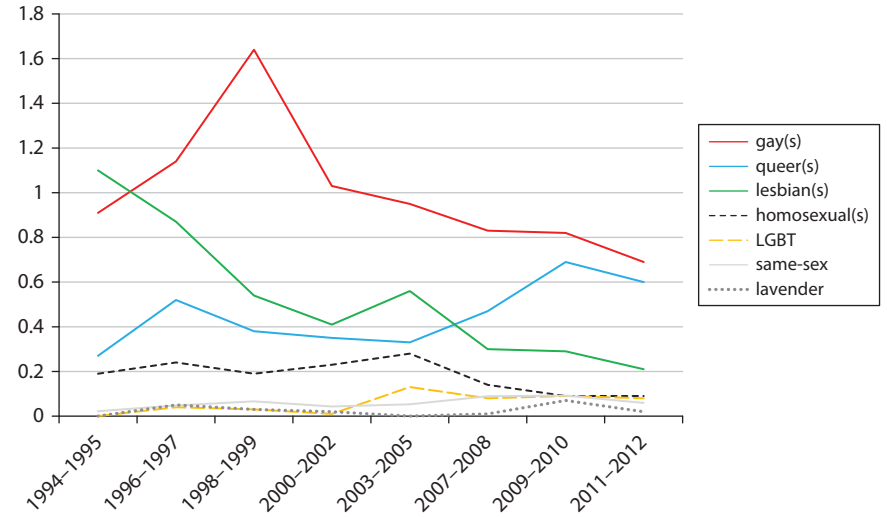

Figure 1. Change over time for sexual identity words (abstracts).

key). *Gay(s)* has generally been the most popular sexual identity word, although this has also been declining since it peaked in the late 1990s.[2] However, *queer(s)* appears to be gaining in precedence, and since 2009 has been at a similar level to *gay(s)*.

Four other terms, *homosexual(s)*, *lavender*, *LGBT* and *same-sex* appear to have had less priority, and *homosexual(s)* has shown a marked decrease since 2005. *LGBT* has shown an increase since 2002 (slightly contributing to decreases in *gay* and *lesbian*), and while it is still relatively infrequent, this increase is enough to have made it a keyword in some of the later time periods. It is notable that *same-sex* and *lavender* have not made much of an impact on conference-goers, especially as *lavender* is in the title of the conference. A question might be raised with regard to whether that word most accurately describes the conference focus, and whether another word like *sexuality* or *queer* might be more effective in terms of promoting the conference, although *lavender* might be viewed by some as a more theoretically neutral term, and in terms of retaining historical consistency, and for its alliterative and thus memorable qualities (e.g. *lavender linguistics*), arguments could be made for maintaining *lavender*.

With three of the words in the Figure 1 being reasonably frequent at various points, it is worth looking at them in more detail, in order to get an impression about their contexts of usage, which may in turn provide clues as to the changing patterns in Figure 1. Table 3 shows the top 10 collocates for the words *gay*, *lesbian* and *queer* across all of the abstracts, using the Mutual Information statistical procedure and only considering collocates if they occurred at least five times, either one place to the left or right of the term being investigated.

Table 3. Collocates of *gay*, *lesbian* and *queer* (abstracts).

	Gay	Lesbian	Queer
1	ex	deaf	theory
2	porn	bi	studies
3	sounding	couples	space
4	Dominican	space	theorists
5	men	bisexual	linguistics
6	Hausa	erotic	youth
7	Russian	identified	theoretical
8	male	desire	people
9	publications	specific	communities
10	immigrant	identity	community

Table 3 shows that *gay* tends to be strongly associated with male identities. In fact, 25% of uses of *gay* in the abstracts occur directly next to the words *men* or *male*. But *gay* is also associated with forms of media (*porn*, *publications*) and language or nationality (*Russian*, *Hausa*, *Dominican*, *sounding*, *immigrant*). There are no gender-related words like *women* which collocate with *lesbian* in the list, but this is not surprising as *lesbian* is generally seen as a hyponym of *woman*. Instead, *lesbian* collocates with words relating to bisexuality, identity, desire and relationships. Finally, *queer* is not gendered at all (either explicitly or implicitly), and it only occurs with a gender-marked word like *male* or *women* 1.5% of the time. Collocates of *queer* like *people*, *community* and *youth* are gender-neutral. Additionally, *queer* is associated with *theory* (and related words *theorists* and *theoretical*), perhaps explaining why the term has become more popular over time, as a gender-neutral term which can relate to various identities or an over-arching theory, it has considerable versatility. The rise of the term may thus signify a gradual acceptance of and/or interest in queer theory in general by people attending the conference. Concordance analyses confirm this, the term is not used by conference-goers in order to criticise it. Queer theory is concerned with deconstructing sexual identity binaries, particularly where one identity is viewed as normative and the other is seen as tabooed or problematic in some way. And while *queer* is sometimes used as a synonym for *gay* or *LGBT*, Cameron and Kulick (2003: 149) argue that anybody has the potential to be queer, including heterosexuals, if their desires or sexual relationships are viewed as non-normative.

The collocation of *lesbian* with *bi* and *bisexual* indicates a grouping tendency in the abstracts — when people refer to lesbians, they are also likely to mention bisexual people. Generally, this practice occurs in 'inclusive' lists like *gay, lesbian, bisexual and transgendered people*. Yet it is telling that *gay* does not have *lesbian* or *bi* as top 10 collocates — this is because *gay* is used in many other contexts where

it is not part of a list of other sexual identity groups, unlike *lesbian*, which can sometimes come across as one of many sexual identities rather than the focus of study in its own right.

Another way that *lesbian* and *gay* are unequally represented relates to gender-marking. As previously mentioned, *lesbian* is already gendered as female although *gay* can theoretically refer to any sex so needs to be marked with a term like *male* or *women* if gender is relevant. As a result, when *lesbian* and *gay* occur together, *lesbian* has a tendency to be used in its noun form, while *gay* is normally an adjectival modifier of *man* or *men*. As Table 4 shows, *gay men* is paired with *lesbians* 42 times across the abstracts. It could be argued that the word *lesbians* is more essentialising than the phrase *gay men* as it reduces women to a single word emphasising their sexuality, while *gay men* posits sexuality as one possible modifier of the male identity under discussion, thus allowing for the possibility of others. Marshall (2004: 8) writes "Using the word gay as a noun implies the identification of a group of people by their sexual preference, as though that, and that alone defined them as people… a tendency which fosters the false assumption of homogeneity".

It is perhaps difficult to resolve this issue: some writers have referred to *gays and lesbians* (or *lesbians and gays*), which occurs 55 times in total in the abstracts. But while this is an equalising strategy, it essentialises both groups so is also open to criticism. Other possibilities like *gay men and women* are very rarely used, perhaps due to their ambiguity — does the writer mean gay men and heterosexual women, or gay men and gay women? Explicitly marking *women* as *lesbian* (*lesbian women*) is also very rare, perhaps because it feels tautological in such contexts.

Table 4. Frequencies of combining male and female gay identities (abstracts).

Term	Abstracts
gays and lesbians	40
lesbians and gays	15
gay men and lesbians	20
lesbians and gay men	22
gays and lesbian women	0
lesbian women and gays	0
gay men and lesbian women	0
lesbian women and gay men	0
gay men and women	1
gay women and men	0
gay women and gay men	0
gay men and gay women	0

There is another inequality in Table 4, relating to male firstness — the male half of the pair appears first overall 61 times while the female half appears first 37 times. These criticisms of the phrase *gay men and lesbians*, as well as the fact that the phrase is rather wordy (yet does not include other identities like *bisexual* or *trans* people) perhaps helps to explain the growing popularity of *LGBT* as a recent key word. *LGBT* manages to bypass some of the issues relating to essentialising some identities more than others. It is also shorter and refers to a greater range of groups, although there is still an issue with the ordering of identities which cannot be avoided, suggesting that the B and T parts could be 'after-thoughts'. In 2013, the term *GSD* (Gender and Sexual Diversity) has been proposed by the group Pink Therapy (http://youtube/7K07qGwk35s) as an even more inclusive alternative, although this term does not appear in any of the conference abstracts.

It is certainly the case that conference-goers appear to have moved away from some of the more obvious essentialising sexual identity nouns. Figure 2 shows relative frequencies for *gays*, *lesbians* and *homosexuals*. While the decline of *lesbians* and *homosexuals* may be mainly or partially attributed to other reasons (such as an increasing dislike of *homosexual*, discussed below), it is notable that *gays* actually increased in frequency up until around 2005, when it then started to fall sharply. Along with the issue of essentialising and homogenising identities, Greenblatt (2011: 221) has criticised the term *gays* as contributing towards lesbian invisibility (due to the fact that *gay* tends to be more strongly associated with male identity), so this could give another reason for the fall of *gay(s)*.

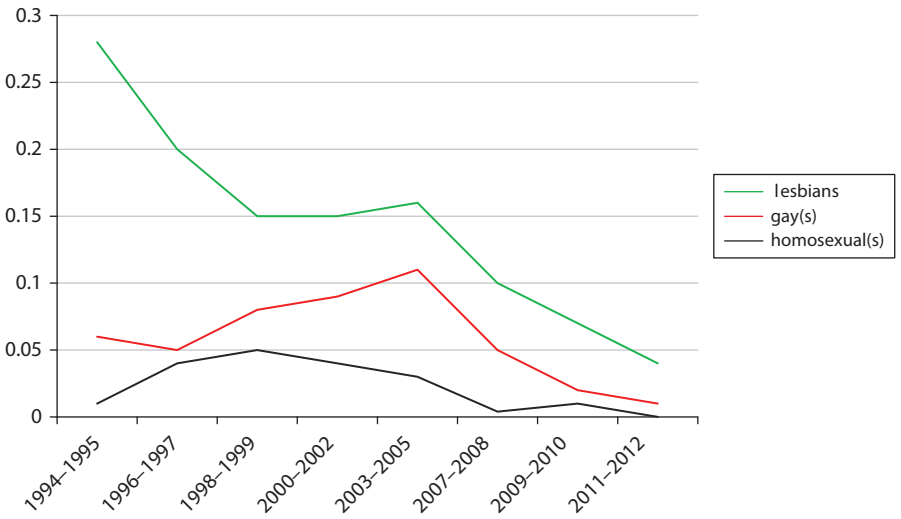

Figure 2. Plural sexuality identity terms (abstracts).

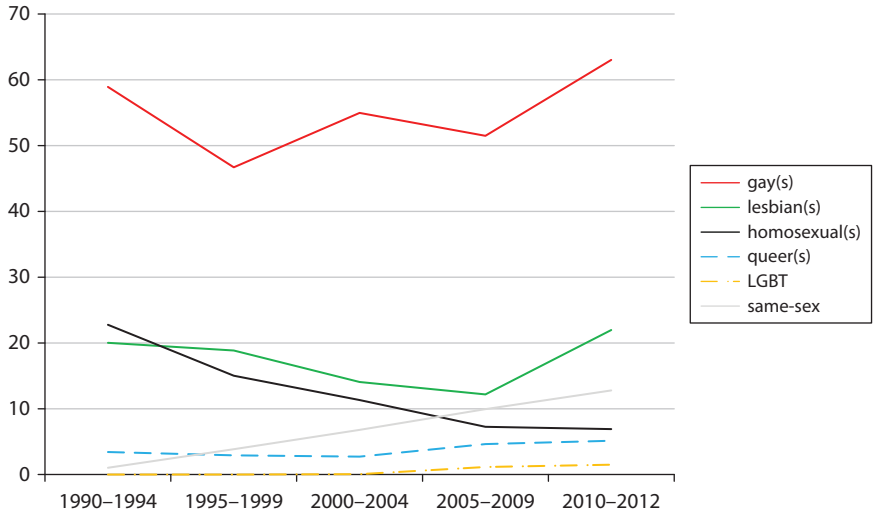

Figure 3. Change over time for sexual identity words (COCA).

How are some of these terms used in the COCA, which stands as a general reference for American English? Figure 3 gives this information, although here the vertical axis shows occurrences per million words (these words are much less frequent in general American English, compared to the conference abstracts).

Figure 3 indicates a clear preference for *gay(s)* which is largely undiminished over time, although it is interesting to note that *same-sex* is steadily rising and that *LGBT* is the least popular term. As with the conference abstracts, *homosexual(s)* is falling, but *queer(s)* shows a smaller increase. In fact, a closer look at *queer(s)* in the COCA reveals numerous cases that are not used in the academic sense but in an older (sometimes pejorative) way, as the following examples attest:

- *It seemed to me the old dimwit had turned* ***queer****.*
- *I seen him chase my friend Jesus into the bathroom, but that was just because my friend made a remark against him, you know, saying, like, you know, he sniffed the air and he said, "I smell* ***queer****," you know.*
- *Good literature is produced by a few* ***queer*** *people in odd corners*
- *The ball had a* ***queer*** *shape*

Unlike the conference-goers, *gay* is still popular in general American English, while *queer* and *LGBT* have yet to make significant inroads. The decline of *homosexual* in both corpora is worth considering further. Table 5 shows the top 10 collocates of *homosexual* and *gay* in the COCA, using the same criteria as described for Table 3.

Table 5. Collocates of *homosexual* and *gay* (COCA).

	Homosexual	Gay
1	copulation	closeted
2	closeted	ambiguously
3	sodomy	openly
4	subculture	legalize
5	stigma	bashing
6	acts	self-identified
7	inclination	marriage
8	tendencies	legalizing
9	conduct	banning
10	orientation	marriages

While *homosexual* tends to collocate with words that focus on sexual behaviour (*copulation*, *sodomy*, *acts*, *conduct*) or uses somewhat euphemistic language that positions homosexuality on a cline (*tendencies*, *inclination*), *gay* is more focussed on politics and rights (*legalize*, *banning*), as well as relationships (*marriage*, *marriages*), indicating a more identity-based conceptualisation of sexuality. Some of the collocates in Table 5 indicate rather negative primings, especially for *homosexual*. For example, in COCA the top 10 collocates of *conduct* are *unsportsmanship*, *disorderly*, *unbecoming*, *unprofessional*, *anticompetitive*, *unethical*, *lewd*, *reprehensible*, *wrongful* and *improper*. It is not fantastical to argue then, that if we hear a phrase like *homosexual conduct* we are likely to infer that this is meant negatively, due to all the other occasions when we have seen *conduct* being used to refer to bad things.

Finally in this section, I report on the frequencies in COCA for *gays*, *lesbians* and *homosexuals*. While there was a general decrease over time of these words, in line with their use in the conference abstracts, in the last time period examined (2010–12), they had all increased slightly. This may represent a 'blip', or it might be indicative of the wide range of voices and political stances that are represented in the COCA. It also perhaps indicates that not everyone is sympathetic to or aware of the criticisms around these words.

4.2 Gender identity words

In this section, I focus on a set of words that relate to gender. First, it is notable that the words *female*, *woman* and *women* do not appear as keywords in Table 2. In fact, *woman* was never key in any of the time periods examined, while *women* was only key in the first period (1994–5). *Female* was only key in two periods (1998–9 and 2011–12). It could be hypothesised then, that women are backgrounded in

the conference, particularly if we consider that *male* was key in every time period, and *men* was key in every time period apart from 1994–5 (although *man* was never key). One important reason for the almost constant popularity of *men* and *male* is due to these words occurring in phrases like *gay men* and *gay male*, while as discussed earlier, we do not see the same pattern for *woman/female*, as *lesbian* normally renders *female/woman* unnecessary. However, it is worth noting that in Figure 1, *lesbian(s)* shows a much more marked reduction than *gay*, and this, combined with the lack of keyness of words relating to women could indicate a declining focus on women in the conference over time, which does not appear to be the case with men. A possible explanation for this could be due to burgeoning lesbian invisibility.

Moving on, I now wish to focus on identities relating to transgender, as there was a clear distinction between *transsexual* which was key in earlier abstracts, and then *transgender* and *trans* which were key later. Figure 4 shows the standardised frequencies of these terms and related ones, in the abstracts. I searched on *trans*, *transsex** and *transgender**, with the asterisk acting as a wildcard, standing for any string of characters. So searching for *transsex** will find *transsexual*, *transsexuals*, *transsexuality*, *transsexualism*, etc.

Figure 4 shows a changing landscape, with all three terms taking precedence at different points. While *transsex** was initially most popular, by 1998–9, *transgender** had taken over, and mostly remained dominant until 2009–10, when the shorter term *trans* gained precedence. This altering focus shows a shift from viewing trans people as being more related to the paradigm of sexuality or sex to that of gender, to then a briefer term. The US National Center for Transgender Equality Guidelines indicate that *trans(gender)* is both a hypernym and a recommended term, stating:

> Transgender: An umbrella term for people whose gender identity, expression or behaviour is different from those typically associated with their assigned sex at birth, including but not limited to transsexuals, cross-dressers, androgynous people, genderqueers, and gender non-conforming people. Transgender is a broad term and is good for non-transgender people to use. "Trans" is shorthand for "transgender." (National Center for Transgender Equality 2009)

The shorter term *trans* perhaps shows a similar process to that of *LGBT*, although rather than being a acronym, *trans* is a clipping, implying a wider range of identities (e.g. *transvestite*, *transsexual*, *transgender*) as there is ambiguity about what can come after it. Both *trans* and *LGBT* can imply inclusivity then. As with *gay* and *lesbian*, there appears to have been a move away from essentialising nouns relating to trans people as *trans(gender)* is used as an adjective in the abstracts,

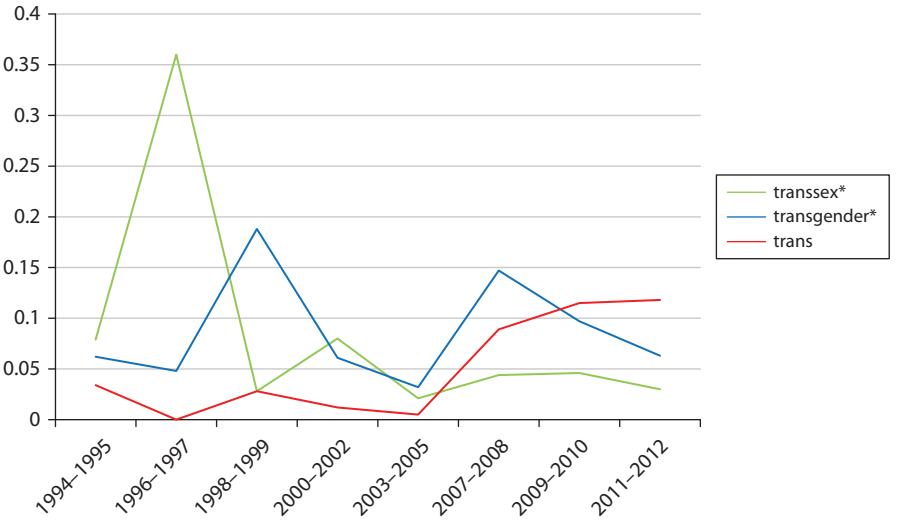

Figure 4. Change over time for words relating to trans identities (abstracts).

whereas *transsexual* can be either a noun or adjective. The two examples below demonstrate changing practices.

- *This paper considers… a male-to-female transsexual who has met with limited success in passing as female… (1997)*
- *Social and biological changes in the voices of trans men (2011)*

I have not included an equivalent graph for COCA here, although note that it generally follows a similar pattern to that of the conference-goers, showing a large decline in *transsex** over time and a rise in *transgender** in its place. However, in general American English, *trans* does not appear to have been adopted in the period under examination, at least in relationship to gender identity. Many of the uses of *trans* in COCA actually refer to *trans fats*.

4.3 Changing themes

In this final analysis section I examine some of the remaining keywords in Table 2. Figure 5 shows relative frequencies of four related terms that can be used to refer to sexuality in a more general way: *sexualit**, *sexual orientation**, *sexual preference** and *sexual identit**.

The figure shows a clear preference for *sexualit** throughout every time period, while *sexual orientation** and *sexual identity** are less frequent and have struggled for second place. The final term, *sexual preference** is very unpopular and has zero occurrences in most of the periods of the conference. *Sexuality*, *sexual*, *identity*

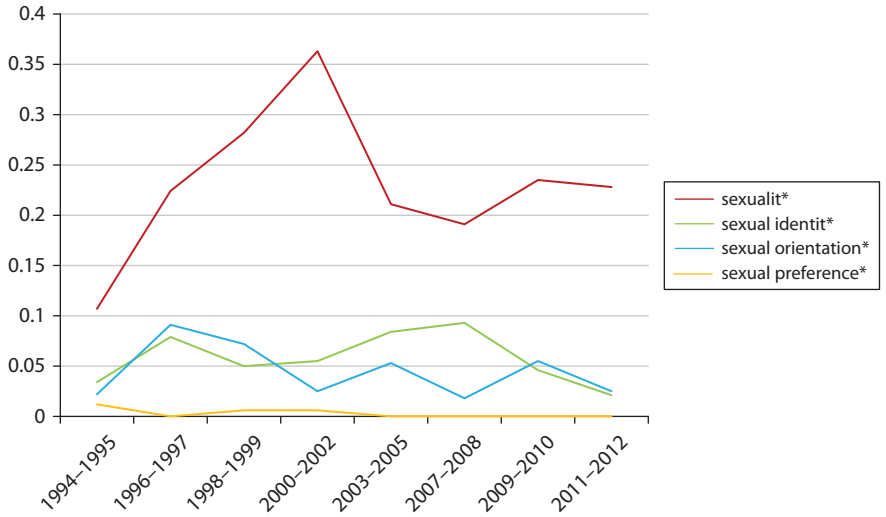

Figure 5. Change over time for *sexuality* and related terms (abstracts).

and *identities* are key across all the time periods in the abstracts corpus, although *orientation* is only key in the earlier periods and *preference* is never key (but was included for the sake of completeness). It is perhaps easy to see why conference-goers would dislike *sexual preference* as the term could imply that people can choose the sorts of people they are attracted to (and thus they can be compelled to change, a key argument of the ex-gay movement which seeks to erase gay people out of existence). But why has *sexual orientation* seen a decline from its peak in 1996–7? Andersen and Taylor (2009: 290) suggest that it implies something deeply rooted in a person. However, if this is the case then it perhaps does not allow for more fluid interpretations of sexuality — perhaps going too far in the opposite direction from *sexual preference*. Weiss (2004: 33) argues that "[w]hen placed under the rubric of 'sexual orientation,' then bisexuality stands out as a failure of orientation or a dual orientation, a product of confusion, promiscuity or indecision". The term *sexuality* avoids these criticisms, as well as being shorter.

Figure 6 shows the equivalent graph for COCA. Again, while *sexualit** is the most common term and *sexual preference** the least, it is worth noting that *sexual orientation** has been increasing at a steady rate in the COCA since 1995–9, a pattern that is not found in the conference abstracts. Additionally, *sexual identity** is almost as unpopular as *sexual preference* in COCA, indicating another difference with the conference abstracts.

The final thread of analysis I wish to pursue begins with the fact that in Table 2, *linguistics* was only key in the earlier years of the conference, perhaps a surprising finding considering that the conference is called *Lavender Languages and*

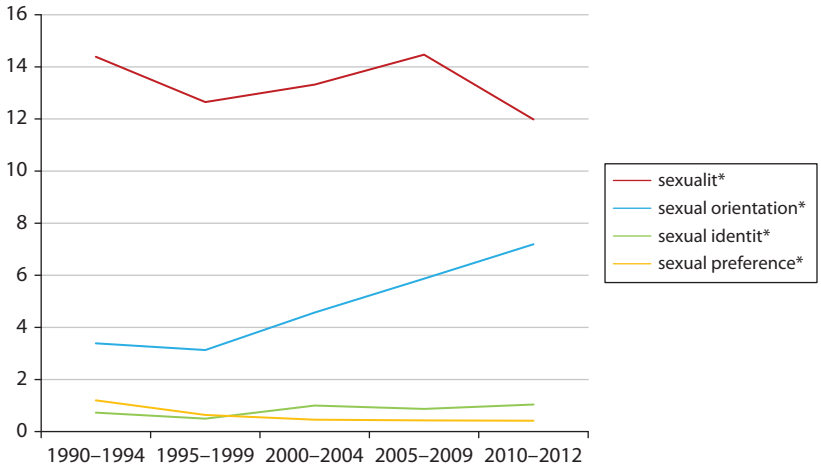

Figure 6. Change over time for *sexuality* and related terms (COCA).

Linguistics. As Figure 7 shows, *linguistics* was never a particularly popular word (in relation to *linguistic* and *language*), although it was still frequent enough in the earlier abstracts to be key. It is interesting to observe that all three terms have actually been declining over time, and the 2010–12 period marks their lowest relative frequencies of any point in the conference.

An examination of the contexts of terms like *language* suggests that a theoretical shift has taken place over time. Prior to 2002, there were a relatively high number of references to LGBT people as speaking a specific form of language:

– *The role of **Gay English** in signalling and maintaining identities (1998–9)*

Figure 7. Change over time for *linguistics* and related terms (abstracts).

- *"How I first discovered gay city" is a widely attested genre of story-telling in **gay men's English** (2000–2)*
- *Can basic concepts of linguistics be taught using examples from **gay language** (1998–9)*
- *The use of certain categories from Argentine **gay speech** (1998–9)*

Collective frequencies of the terms in bold peaked in 1998–9 and have declined sharply since then, so that there were no occurrences of them in 2011–12. Does this indicate then that the conference has rejected the concept of gay people as having their own language? The following examples give cases of *language* in 2011–12:

- *language in the queer community*
- *Russian gay men must be fluent in the language of their national gay community*
- *the use of a certain language has served the queer community*
- *How is language used by these men to resist a gay identity*
- *language associated with queer identities*

In more recent times, I would suggest that conference-goers have not given up on the idea that gay people use language in specific ways, although they have shied away from labelling this as *gay language* or *gay English*. Instead, they tend to use more cautious and qualified phrasing which focuses on language *use* or language *associated with* an identity. However, it should also be noted that references to *language* and *linguistic(s)* have generally declined over time, indicating that even these more careful references to language usage tend to be lower overall in the later periods of the conference.

A question was raised at this point to do with what other contexts *language* could be used in, if conference-goers are less likely to use terms like *gay language*. Related to the rise of the term *queer* and *homophobia/homophobic*, I wondered whether there had been increased focus over time on language used *about* LGBT people. Figure 8 shows all of the occurrences of *language* in the abstracts (reduced to four time periods rather than eight to make comparison easier). I examined concordance lines and noted when *language* was used to refer to use by LGBT people and cases where language was about LGBT people. I had expected to find that over time, there would be a decline in references to language used by gay people and more emphasis on the language used about them, linking to a shift towards studies which look at representations of LGBT people. However, this hypothesis was only partially borne out by Figure 8. While there does appear to have been a decline in references to language use by LGBT people, there is only a smallish increase in the language about LGBT people occurring in 2003–8, but such cases are barely present in the 2009–12 period.

A second bar chart (Figure 9) helps to explain this unexpected pattern.

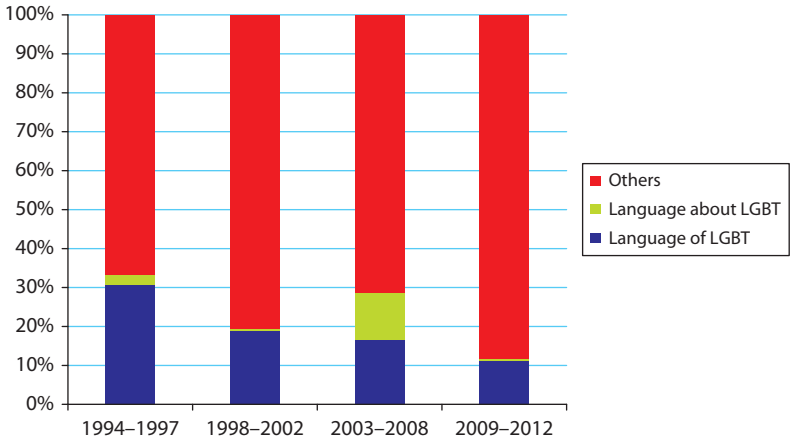

Figure 8. Uses of *language* over time (abstracts).

Figure 9, focuses on a related word: *discourse*, which is always key (as shown in Table 2). Again, I looked at concordances, noting when people referred to discourse used by LGBT people and cases where discourse was about LGBT people. Here, the former decreased over time, while the latter showed more of a relative increase (at least compared to *language*). I would suggest that *discourse* is perhaps doing some of the work that *language* used to do in the conference, although *discourse* appears to be used in a more versatile (or in some cases vague) way, referring both to ideologies or representations of (LGBT) people in a more general sense, as well as implying the actual language that is used in order to index such ideologies:

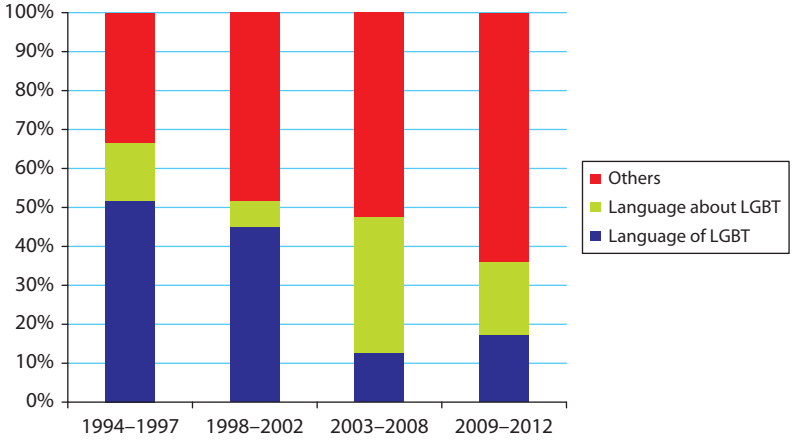

Figure 9. Uses of *discourse* over time (abstracts).

- *fear, stereotyping, and ageism in the discourse of LGTB colleagues. (2010)*
- *determining the nationally shared knowledge that is used in Russian gay discourse (2011)*
- *Newspaper discourse surrounding the LGBT community (2011)*
- *The discourse of gay marriage in Argentina (2011)*
- *LGBTQ discourses in School Setting (2012)*

Some of the later keywords in the abstracts indicate a focus on topics related to discourse: *homophobia*, *homophobic*, *normative* and *discursive*, although not *representations* which has declined.[3] Issues relating to discrimination and prejudice faced by LGBT people, as well as the ways that heterosexuality is assumed to be ubiquitous and/or 'normal' are thus increasingly popular topics of conference papers in more recent years. It can thus be seen then, that the general picture in terms of focus is to do with a move away from the idea that gay people have their own 'language' and a move towards studies which carry out discourse analyses, particularly in order to comment on issues relating to LGBT equality. Perhaps what is surprising then, is that the text on the conference website which describes its focus does not include the word *discourse*, even though this word has been key for every time period in the conference, and is arguably even more important than it used to be, due to what appears to be a declining focus on linguistics. The following examples show some of the more typical foci in later periods of the conference, focussing on homophobic and (hetero)normative discourse or discursive practices and queering normative texts. Such foci are in keeping with the goals of queer theory and are perhaps unsurprising, considering the growing popularity of *queer* over time.

- *Homophobia and the discourse of gay rights (2010)*
- *…act of taking a normative text and exploring the ways in which it can be queered. (2010)*
- *…the analysis of homophobic discourse must consider the larger playing field of racist and sexist discourse (2011)*
- *This presentation focuses on unpacking heteronormative Discursive practices in school settings (2011)*
- *Tasked with moving away from (hetero)normative discourse… (2012)*

It is perhaps surprising to see a decline in the word *culture* over time as this was an early keyword but not key in the later periods, although this could be related to the same issue to do with critiques of 'gay language'. In the abstracts corpus, there are 13 references to *gay culture*, 9 to *gay-male culture*, 4 to *lesbian culture*, 4 to *queer culture* and 3 to *gay youth culture*. Of these 32 references then, 21 (66%) occur in the period 1994–2002 rather than 2003–12. Again, references to gay or lesbian

culture may be critiqued as being separating and homogenising, ignoring differences within particular sexual identity groups, which may explain the decline of *culture* in the abstracts.

Finally, the term *desire* has also declined, being key in three of the earlier periods of the corpus abstracts but not key at all in the later four periods. The shift away from *desire* seems to have occurred around 2002–3, and in 2003, a couple of papers are critical of the ability of a desire-based framework to fully explain sexuality:

- *desire is not always relevant to the material ways in which sexuality operates, and even when it is, it is always mediated in some way by identity*
- *This session responds to recent discussions regarding language and sexuality which urge that researchers reduce the contexts, content, styles and attitudes expressed through LGBTQ language to a single topic: language about desire… recent articles on the subject are most closely associated with this reductive argument*

It is therefore noticeable that references to *desire* appear less often in the conference after this point, indicating perhaps that attempts to position desire as a central concept (perhaps even *the* central concept) for research in Language and Sexuality were unsuccessful, while the continuing keyness of the terms *identity* and *identities* indicate how this concept has held on, despite the shift away from considering identity in relation to concepts like *gay language*.

5. Conclusion

In this final section, I summarise and evaluate some of the emerging themes from conference abstracts, then compare the changes in the conference with what has been happening in wider (American) society. However, I begin with a methodological consideration.

While I find the keywords procedure to be useful in helping to provide a focus for analysis, as well as reducing research bias, one point of criticism is that the approach I have taken only gives keywords which indicate presence rather than absence. To put it another way, certain concepts may be so ingrained or taken-for-granted amongst conference-goers, that they are simply never referred to very often, and if they are equally (in)frequent in a reference corpus, they will never emerge as key. For example, conference abstracts are written and presented in English and a good many of the papers are focussed around English texts or data. When a paper analyses a language other than English, this is usually made explicit, explaining why *French* was occasionally a keyword. However, we might expect

English to be such an assumed norm that people do not feel the need to mention that their data is in that language. In fact though, *English* was a keyword in two time periods, particularly due to its inclusion in the phrase *gay English*. One concept which does appear to be sometimes taken for granted is America. As noted earlier, the conference has always taken place in America, many of its delegates are living in the US or are American citizens, and while there are references to *America** in the abstracts, they do not occur enough to result in keyness (particularly when compared against a reference corpus containing other American texts). Additionally, when conference-goers do use the term *American*, they often refer to *African Americans*, *black Americans*, *Asian Americans* or *Latin Americans* rather than Americans per se. The generic American subject is unmarked for conference-goers normally.

Another word which is never key is *white*, although the word does occasionally occur in abstracts particularly in discussions of ethnicity, in relation to other ethnic groups, while *black* is a keyword in the 1994–5 abstracts. There are twice as many references to *black* than *white* across the corpus, and *black* tends to be used in marked cases, when a conference presenter is focussing on say, the language use or representations of black gay men as a group. Again, we do not tend to find such marking of white gay men, they are generally just gay men in other words.

Therefore, while the keywords technique can help to pick out changes, unless a range of different reference corpora are used or the method is combined with others, it can be difficult to identify mostly 'unsaid' concepts that are infrequent because they are often taken for granted.

Let us move on to what the analysis indicates. To summarise, it appears that conference-goers have evolved more careful language practices around words that relate to sexual and gender identity. There has been a move away from separating, homogenising identity terms like *gay* and *lesbian* (and especially plural noun versions of these terms), and a move towards more inclusive terms like *queer* and *LGBT*. Related to this, there has also been a move away from the idea of LGBT people as having their own language (or culture) and greater focus on critiquing (hetero)normative discourses.

Comparing the language of conference presenters with language use in general American English, there are some notable differences: the conference is increasingly centring on *queer*, which threatens to overtake *gay* in terms of frequency. But in American English, *gay* continues to increase and *queer* barely registers (and is sometimes pejorative). Another difference is in use of *sexual orientation* — a term which is growing in popularity in general American English but less common among conference attendees. An opposite pattern is found with *trans*, however, which is yet to be taken up in general American English. These changes in terminology, especially relating to sexual and gender identities, indicate that both

academics working in the field of Language and Sexuality, and non-academics are still in the process of critiquing and debating the most appropriate ways to name different identity groups. Rather than viewing this process pessimistically as being an example of Pinker's (1994) 'euphemism treadmill', where new 'politically correct' terms are continuously and somewhat redundantly suggested in order to replace older ones which attract negative meanings, I would instead argue that these changes in terminology reflect an increasingly sophisticated understanding of sexual and gender identities. A phrase which emphasises identity like *gay men* and *lesbians* is preferable to *homosexuals* which focuses on sexual behaviour. However, terms like *LGBT* and *queer* are more inclusive and avoid the discrepancy between *lesbians* as a noun and *gay* as an adjective. These alternative wordings are not simply 'politically correct' synonyms of one another, but hold different, if related, meanings, suggesting increasing thoughtfulness and sophistication, and reflecting the continually changing social status of the subjects of such identity labels.

A problem with evolving terminology is that it requires people to (want to) adapt their language. People can sometimes find it difficult to drop older terms, and at any given time in a society there will simultaneously exist several diachronic 'layers' of terminology. Comparing the figures for the abstracts with those for COCA, I suspect that presenters at the Lavender Languages and Linguistics conference represent the forefront of thinking around these issues, and it would be interesting to see whether terms like *queer* and *trans* become more popular in later editions of the COCA. However, a danger of using a 'discursively advanced' form of language is that sections of the population may feel excluded, confused or jaded by the changing linguistic landscape. On several occasions I have realised that my understanding of *queer* is different from that of non-academics which has resulted in temporary communication failure. While we may want to lead change from the front, we also need to be able to engage with a wide range of communities, and we can most effectively do that if we take care with our use of academic jargon that has yet to enter into the public domain. Yet this presents a dilemma: we should not 'dumb down' or abandon academic terminology like *heteronormative* and *discursive* as these concepts are theoretically and analytically useful, but we need to be able to get those concepts across to a wider audience without our own language appearing annoyingly impenetrable. My advice would be to use such concepts in non-academic writing sparingly and with sufficient exemplification. An area for further research would be to examine how non-academics orient to some of the more arcane-looking terminology from the field of Language and Sexuality and whether lessons can be learnt from analysing the ways that a wide range of people engage with such terms.

While I am happy that certain language practices are in the process of being abandoned (*gays*, *homosexual*), I am somewhat concerned about the shift away from *linguistics* in the conference abstracts. The focus on *discourse* is welcome, but I feel that discourse and language are complementary rather than incompatible, and I would hope that linguistics and language continue to play a central role in the conference for years to come. Similarly, I wonder whether the critiques of *desire* a decade ago have perhaps led to conference-goers moving too far away from this concept. Stepping outside the corpus analysis of abstracts for a moment, those involved in the field around the turn of the century will probably recall a debate about central theoretical underpinnings and foci, particular the concepts of desire and identity (see Cameron & Kulick 2003, Eckert 2002, Kulick 2000). I did not take part in this debate, and while my feeling is that desire cannot explain everything, it is an important concept and deserves more emphasis than it currently seems to have. While desire was suggested as an alternative focus to research which reified the concept of 'gay language' and shifted attention away from identity, it appears that in later years of the conference, there has been a shift away from both desire and 'gay language' but not identity. Another 'winner', theoretically at least, has been queer theory and its bedfellow normative discourse. As battles for LGBT equality are gradually won in rich western countries, it will be interesting to see whether the focus will change again, perhaps moving on from analysis of homophobia and LGBT people, instead encompassing other configurations of sexual identity, desire and behaviour. There could (and I argue, should) also be increased attention on non-Western contexts in order to challenge what sometimes appears to be an underlying assumption that *gay* means American white man. The next 20 years promise to be as equally fascinating as those that have gone before them.

Notes

1. The research presented in this paper was supported by the ESRC Centre for Corpus Approaches to Social Science, ESRC grant reference ES/K002155/1.

2. Since the 2000s, *gay* has had a pejorative meaning in informal American English (particularly among younger people), although conference-goers do not use it in this way, and it is difficult to assess the extent to which this meaning could have influenced usage of the word for conference-goers.

3. While *representation* has declined, *presentation* is a later keyword. However, the two words are not directly related. *Presentation* tends to be used in contexts like 'In this presentation I will…' rather than as a synonym for *representation*. I would argue that in earlier years of the conference, the words *language*, *representation* and *discourse* were sometimes used interchangeably or at least had overlapping meanings, perhaps reflecting the different sorts of terminology used

by conference-goers from a wide range of disciplines. Later in the conference, *discourse* seems to have become more popular, possibly indicating a move towards consensus.

References

Andersen, Margaret L. & Taylor, Howard F. 2009. *Sociology: The Essentials*. Belmont, CA: Thomson Wadsworth.

Anthony, Laurence. 2011. *AntConc (Version 3.2.2) [Computer Software]*. Tokyo: Waseda University. Available from <http://www.antlab.sci.waseda.ac.jp/>

Baker, Paul. 2010. Will Ms ever be as frequent as Mr? A corpus-based comparison of gendered terms across four diachronic corpora of British English. *Gender and Language* 4(1): 125–129.

Baker, Paul, Gabrielatos, Costas & McEnery, Tony. 2013. *Discourse Analysis and Media Bias: The Representation of Islam in the British Press*. Cambridge: Cambridge University Press.

Cameron, Deborah. 2005. Language, gender and sexuality: Current issues and new directions. *Applied Linguistics* 26(4): 482–502.

Cameron, Deborah & Kulick, Don. 2003. *Language and Sexuality*. Cambridge: Cambridge University Press.

Campbell-Kibler, Kathryn, Podesva, Robert J., Roberts, Sara J. & Wong, Andrew (eds). 2002. *Language and Sexuality: Contesting Meaning in Theory and Practice*. Stanford, CA: CSLI Publications.

Clark, Caroline. 2010. Evidence of evidentiality in the quality press 1993 and 2005. *Corpora* 5(2): 139–160.

Culpeper, Jonathan. 2009. Keyness: Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's Romeo and Juliet. *International Journal of Corpus Linguistics* 14(1): 29–59.

Duguid, Alison. 2010. Newspaper discourse informalisation: A diachronic analysis from keywords. *Corpora* 5(2): 109–138.

Eckert, Penelope. 2002. Demystifying sexuality and desire. In *Language and Sexuality: Contesting Meaning in Theory and Practice*, Kathryn Campbell-Kibler, Robert J. Podesva, Sarah J. Roberts & Andrew Wong (eds), 99–110. Stanford, CA: CSLI Publications.

Greenblatt, Ellen. 2011. *Serving LGBTIQ Library and Archives Users*. North Carolina: MacFarland.

Holmes, Janet & Sigley, Robert. 2001. What's a word like *girl* doing in a place like this? In *New Frontiers of Corpus Linguistics*, Pam Peters, Peter C. Collins & Adam Smith (eds), 247–263. Amsterdam: Rodopi.

Johnson, Jane. 2012. What are these corpus linguists talking about? An MD-CADS content analysis of the IJCL. Paper presented at the *Corpus Assisted Discourse Analysis Conference*, Bologna, Italy, 14 September 2012.

Kulick, Don. 2000. Gay and lesbian language. *Annual Review of Anthropology* 29: 243–285.

Lavender Languages. 2013. Background & focus. <http://www.american.edu/cas/anthropology/lavender-languages/background.cfm> (May 18, 2013)

Leech, Geoffrey. 2003. Modals on the move: The English modal auxiliaries 1961–1992. In *Modality in Contemporary English*, Roberta Facchinetti, Manfred Krug & Frank Palmer (eds), 223–240. Berlin: Mouton de Gruyter.

Leech, Geoffrey. 2011. The modals ARE declining: Reply to Neil Millar's 'Modal verbs in TIME: Frequency changes 1923–2006.' *International Journal of Corpus Linguistics* 14(2): (2009), 191–220. *International Journal of Corpus Linguistics* 18: 547–564.

Leech, Geoffrey & Fallon, Roger. 1992. Computer corpora — what do they tell us about culture? *ICAME Journal* 16: 29–50.

Marchi, Anna. 2010. 'The moral in the story': A diachronic investigation of lexicalised morality in the UK press. *Corpora* 5(2): 161–189.

Marshall, Sue. 2004. *Difference and Discrimination in Psychotherapy and Counselling*. London: Sage.

Millar, Neil. 2009. Modal verbs in TIME: Frequency changes 1923–2006. *International Journal of Corpus Linguistics* 14(2): 191–220.

Motschenbacher, Heiko. 2012. *An Interdisciplinary Bibliography on Language, Gender and Sexuality (2000–2011)*. Amsterdam: John Benjamins.

National Center for Transgender Equality. 2009. Transgender terminology. <http://transequality.org/Resources/NCTE_TransTerminology.pdf> (May 18, 2013)

Oakes, Michael. 2003. Contrasts between US and British English in the 1990s. In *Research and Scholarship in Integration Processes. Poland — USA — EU*, Elżbieta H. Oleksy & Barbara Lewandowska-Tomaszczyk (eds), 213–222. Łódź: Łódź University Press.

Pinker, Steven. 1994. *The Language Instinct: How the Mind Creates Language*. London: Penguin.

Potts, Amanda & Baker, Paul. 2012. Does semantic tagging identify cultural change in British and American English? *International Journal of Corpus Linguistics* 17(3): 295–324.

Scott, Mike. 2008. *WordSmith Tools (Version 5) [Computer Software]*. Liverpool: Lexical Analysis Software.

Sigley, Robert & Holmes, Janet. 2002. Looking at *girls* in corpora of English. *Journal of English Linguistics* 30(2): 138–57.

Smith, Nicolas. 2002. Ever moving on? The progressive in recent British English. In *New Frontiers of Corpus Research*, Pam Peters, Peter C. Collins & Adam Smith (eds), 317–330. Amsterdam: Rodopi.

Taylor, Charlotte. 2010. Science in the news: A diachronic perspective. *Corpora* 5(2): 221–250.

Weiss, Jillian Todd. 2004. GL vs. BT: The archaeology of biphobia and transphobia within the U.S. gay and lesbian community. *Journal of Bisexuality* 3(3–4): 25–55.

Author's address

Paul Baker
Department of Linguistics and English Language
County South, Lancaster University
Bailrigg, Lancaster, LA1 4YL
United Kingdom

j.p.baker@lancaster.ac.uk