McEnery, Anthony, and Helen Baker. "Looking Back, Looking Forwards." *Corpus Linguistics and 17th-Century Prostitution: Computational Linguistics and History*. London: Bloomsbury Academic, 2017. 199–205. Corpus and Discourse. *Bloomsbury Collections*. Web. 26 Apr. 2023. <<http://dx.doi.org/10.5040/9781474295062.0011>>.

Downloaded from Bloomsbury Collections, www.bloomsburycollections.com, 26 April 2023, 12:17 UTC.

Copyright © Anthony McEnery and Helen Baker 2017. Released under a CC BY-NC-ND licence (<https://creativecommons.org/licenses/by-nc-nd/3.0/>). You may share this work for non-commercial purposes only, provided you give attribution to the copyright holder and the publisher, and provide a link to the Creative Commons licence.

Looking Back, Looking Forwards

We thought we would end this book by reflecting back on how we carried out our study and the main points arising from it.

Let us begin by considering the choices we made when we decided how to do this study. In addition to selecting the methods that we have used, we decided, quite consciously, for Chapters 2 to 6 to work independently initially. The first drafts of Chapters 2 and 3 were written independently of Chapters 5 and 6. Later, when the drafts of the chapters were complete we compared them and drew the links between them, which we highlighted mainly in Chapters 5 and 6. This point is crucial. The linguist did not look for things the historian told them to. Nor did the historian fish for evidence to explain or problematize what the linguist had found. Rather, we worked independently and found that there was much discovered by the linguist that added weight and credence to what historians have said and found. As well as adding weight to claims in historical research, the equation reverses nicely too – the work of the historian serves well to provide historical explanations of much that we see in our corpus. So the main point to be gleaned from Chapters 2 to 6 is that of a useful bidirectional symbiosis. Corpus linguists can provide evidence and frame an issue. Historians can bring their own evidence, from close reading, for example, or through methods proposed by conceptual historians. In doing so some new methodological synergies may occur – corpus linguistics could be a way of countering the criticism of works such as those by Koselleck¹ which is heavily based on continental philosophy and has consequently been described as ‘impressionistic’.² Similarly, work by the cultural researcher Raymond Williams, while it appears to offer an approach to the study of words in the past, relies heavily on impressionistic analyses and an uncritical adoption of dictionary evidence, an unwise decision given the results of the critical exploration of the utility of contemporary and historical dictionaries for exploring the language of the seventeenth century as carried out in this book.³ Nonetheless, keyword analyses may prove to be a type of evidence that

historians could bring, and refine, in the context of a corpus-based approach to words in the past.

It is undoubtedly the case in our view, however, that the historian has a key role to play in working with the linguist to help provide explanations within the framing of the data, by corpus linguistics and other methods. Those explanations contextualize what is observed in the data in the society behind the production of the texts in the corpus.

Approaching things this way had another merit. In advance of the analysis we considered various issues which could, at that stage, have led us into what would have probably been a lengthy and sterile debate. For example, when we initially discussed the idea for this book, the historian had some reservations. Three examples will suffice to show the nature of the reservations. First, there was the question of what any quantification would actually show – it was clear that the corpus could quantify, but what was the purpose of that? Secondly, while one word to refer to prostitutes may be the word of choice to refer to them at the beginning of the century and another at the end, how was that of interest? If the words meant the same the change has little significance. Finally, if we accept quantification is important, surely it is the frequency of wordforms that is most important – why bother with collocation and other techniques? By sidestepping this debate and actually working on a problem, we were able to much more productively debate these issues after we had done our draft analyses. Rather than speculate on the basis of what we thought the data may or may not show, our debate became focused on what we had actually done. Hence, we could revisit the questions and test them against our findings – we saw that quantification did help as the analyses in Chapters 5 and 6 have shown. For example, quantification is the key to finding consistent, initiating, terminating and transient collocates. Without being able to quantify collocation, it would be very difficult to make these observations systematically.

We also demonstrated that words like *harlot* and *whore* are near synonyms, not pure synonyms – when one becomes less frequent and the other becomes more frequent this can index a change in attitude accordingly. Also, through collocation in particular we have shown, as we do to some extent in Chapter 1 anyway, why a wordform-based analysis is inadequate. Words have multiple meanings in use and that meaning changes over time. Collocation can help us address those issues, the frequency of a wordform like *prostitute* on its own cannot. Consequently, rather than worrying about what would happen when we did our analysis, we did our analyses, integrated them and then were able to identify and dismiss non-issues, which may have looked substantial in advance,

swiftly. Yet we also found that some of the non-issues pointed to areas of very real agreement between historical research and linguistic research. For example, the issue of meaning change over time is shared ground between historians and linguists. Historians may talk of conceptual history while linguists may prefer to talk of semantic change, but both are basically working with the same issue, though perhaps with different emphases and purposes in mind.

If these findings were the only ones to be reported one might say that, perhaps, by working separately we were being a little precious. The linguist might have worked more swiftly by working under the direction of the historian. If the sole role for the corpus linguist is to provide supporting confirmatory evidence and an overall framing, then they can sensibly be subordinated to the historian. However, this is not the case. Two broad types of evidence support this conclusion.

First, there are events which have been given salience in historical accounts of prostitution in the seventeenth century that are not at all salient in the public discourse of the time. It may well be that these were indeed salient events, but they were not written about very much if at all. A good example of this is the Apprentice Riots, as discussed in Chapter 3, Section 3.3. They do not figure in any way in the analyses presented in Chapters 5 and 6. No findings in those chapters seem to need the Apprentice Riots as a background explanation. This stands in contrast to other findings in those chapters that did draw on the evidence from Chapters 2 and 3. Whether or not the Apprentice Riots were salient in forming public views on prostitution the linguist cannot say – but the linguist can say that there seems scant to no evidence of these riots having entered public discourse in any significant way, either directly or indirectly. It may be the case, of course, that something that occurs in society does not enter public discourse clearly until sometime after the event. A modern example of this might be the ship the *Empire Windrush*. This ship brought one of the first large groups of West Indian migrants to the UK in 1948. How did this impact on public discourse? To explore that we looked at the Times Online – for the first ten years after the arrival of the West Indian migrants, the *Windrush* is mentioned, but largely in its role as a troop carrier for the British Army. In the four decades following that, the ship is mentioned twice. However, as the importance of what occurred when the *Windrush* docked enters the public consciousness and the role of the ship as an icon of a major change in British society is established, in the period 1997–2006 there are forty-one mentions of the ship in the newspaper, with its role in public discourse established. It takes nearly fifty years from the docking of the ship to a point where it becomes a

shorthand for discussing the Afro-Caribbean migration to the UK. While many of the examples given in Chapters 5 and 6 do show that some effects on public discourse can be rapid, we should always be alert to the possibility that there may be a very significant lag between an event and its impact on public discourse being realized. While we think that it is unlikely that the Apprentice Riots did, in fact, have an impact on discourse beyond the range of the dataset we used, we should accept that one of the possible explanations for a failure to observe an event in public discourse is that the impact of the event is delayed. Finally, we should note another possibility. We are looking at public discourse. It may be that, if we were able to look at private discourse in the period we may have found somewhat different patterns. When looking at public discourse in the period we are looking at what those, typically powerful males, with access to printing presses chose to print. Even that is filtered through the censorship regime, both formal and informal, of the period. So while public discourse has the merit of being accessible to us, even centuries later, we should always remember that the totality of discourse in the century is inaccessible to us. What we can see is the extant public discourse that has survived the best part of four centuries. While we may make the simplifying assumption that what remains is typical of discourse as a whole at the time, we must remain sensitive to the possibility that this may not be the case.

The second type of evidence is more categorical and clear-cut. There were occasions in Chapters 5 and 6 where a claim made by an historian was explored and doubt was cast upon it. For example, we explored the observation that *PROSTITUTE* seemed to be a verb only in the century – this was disproven.⁴ Similarly, we explored the claim that *WHORE* did not denote a person who had sex for gain. While we found examples where this may have been true, we also observed many cases where it was false, and saw how being paid for favours given seems to form a stable part of the meaning of the word *WHORE*. So there are occasions where observations based on hand and eye studies may be challenged when we engage with a much larger dataset using corpus techniques.

Note that the linguist did not mine for evidence to refute claims – the linguist undertook an independent analysis within which, when the findings were compared to the work undertaken by historians, some claims by historians were problematized. One could, of course, approach the study quite differently, actively seeking to falsify. However, this was not our goal. What we wanted to do was compare two ways of working and show how, quite naturally, they could integrate. In doing so, at times, the claims do not match and then evidence has to be weighed and a finding made. However, we were struck by how often an

investigation of public discourse did mesh well with the established research undertaken by historians on the topic.

Some of the unexpected findings, however, shed light on seventeenth-century thought. Academics and social commentators currently debate whether prostitutes are empowered or exploited women.⁵ Yet it appears that this dichotomy was alive in discourse four centuries ago as shown in Chapter 5. The collocates *play*, which suggests a woman with agency, and *make*, which gives the impression of a woman forced against her will, represent those opposites. As shown in Chapter 5, it is the empowered whore which dominated seventeenth-century discourse. Yet other elements of the corpus analysis pose questions which are not easily answered – why, for instance, in terms of collocates of *whore*, did *keep* not re-enter discourse as a collocate until the 1680s, and why did the *pity* collocate terminate in the 1680s just when historians have suggested that popular feelings of compassion directed towards prostitutes were starting to emerge? These results are not easily explained. We know that the phenomenon of keeping mistresses was very popular at court from the Restoration of the monarchy in 1660 but it is possible that *keep* did not re-enter popular discourse because the sexual mores of the general public lagged behind those at court. This result, therefore, possibly illuminates a much wider aspect of early modern history than expected. In terms of the collocate *pity* disappearing in the 1680s, we have posited that the moral reform movement that emerged at the end of the century may have temporarily stifled discourses of sympathy directed towards prostitutes. These ideas require further research by historians.

Yet there are also clear challenges for linguists. The analyses presented in Chapter 4 show, quite clearly, we believe, how revolutionary a resource like EEBO can be for the lexicography of the period. The corpus enriched our view of the words used to refer to prostitution and other forms of transgressive sex in the period. Existing dictionaries lacked much information that the corpus can provide, from evidence for relevant new entries to important frequency information. Similarly, our corpus showed a much more nuanced view of the preference for the use of various words over time and how the meaning associated with wordforms changed over time. All of this was information which is provided only relatively crudely for this period in the best dictionaries available.

We also had the opportunity to reflect on the lexicography of the seventeenth century and concluded that this is largely of use for looking at hierarchies in word meanings and at near synonymy. As native speaker intuition evidence from the period for these things, the dictionaries of the period are helpful for

our purpose. However, to determine word meaning they are significantly worse than modern dictionaries and provide nothing of the depth or nuance of analysis that a large corpus from the period can achieve.

Near synonymy remained an issue in Chapters 5 and 6 and a major burden for the analysis across those periods is to try to prise the meanings of the words which are near synonyms, and collocates, of one another apart. Such an analysis will only ever succeed in this role by degrees precisely because the words are near synonyms – they share meaning. They cannot be wholly prised apart. However, the distinctive features of the words can be identified, as we showed in Chapter 6. This can be very helpful for improving the entries in historical thesauri, for example. One resource we have not mentioned in this book so far, but which is relevant here, is the *Oxford Historical Thesaurus*.⁶ This very helpfully identifies the type of ambiguity and range in meaning that we have discussed – so, for example, WHORE is classified both as ‘a prostitute’ and as an ‘unchaste or loose woman’. Note that the historical thesaurus entry for this word generally fits well with the analysis in Table 5.3, especially if we see the unchaste or loose woman category as being linked to an intent to wound or offend. However, there are words which, as Chapter 6 has shown, also refer clearly to prostitutes but which are not classified in the thesaurus under prostitution – for example, *harlot* and *strumpet*. As a noun STRUMPET is classified as an ‘unchaste or loose woman’ but not as ‘prostitute’ though it appears in other classes which define it as a verb ‘to bring to condition of whore’ and as an adjective ‘relating to or of the nature of prostitution’. HARLOT is even more interesting as many of its classes relate to males – something not attested as salient in our analysis of the word; like *strumpet* it appears in ‘unchaste or loose woman’ but not in the ‘prostitute’ class. So while the *Oxford Historical Thesaurus* represents the best historical thesaurus currently available, this brief investigation explains why it was only of passing use to our study. As was the case with the dictionaries explored in Chapter 4, a large corpus like EEBO will provide dictionary and thesaurus makers with the data which will allow them to greatly improve these resources, as they were able to do for present-day English when the large corpora of late-twentieth-century English became available. Hence our observations about the dictionaries and thesaurus are intended as proof of need for work in the area and as a spur to more work to improve those resources. The benefits that corpus data brought to lexicography of present-day English will be as relevant to Early Modern English now the data is available.

The work that lies ahead for linguists and historians in exploiting, at times together, at times separately, a resource like EEBO is vast. Returning finally to

observations made in Chapter 1, we hope that new techniques for exploring such data will become available. Indeed, as they speed and make more efficient our analyses they will be welcome. Yet we do not see a context at present, or in the near future, where computational techniques will negate the need for close reading or analysis and explanation by subject experts. Consequently, we look forward to a future in which tools continue to reduce the problem space that linguists and historians work with – focusing in on key texts or linguistic relations that help the analyst to rapidly investigate and account for large volumes of data. We do not believe in a future in which push-button analyses performed and interpreted by researchers with no expertise but the ability to programme and apply complex mathematical models are the norm. We look forward instead to a future in which interdisciplinary teams, using tools which undertake tasks that they find interesting and relevant from their disciplinary perspective and which offer explanatory, not simply descriptive, power dominate. The tools on the shelf today make that a viable prospect. Working with programmers and mathematicians to get the right tools on the shelf tomorrow strikes us as key.