News as Changing Texts

News as Changing Texts: Corpora, Methodologies and Analysis

By

Roberta Facchinetti, Nicholas Brownlees,
Birte Bös and Udo Fries

CAMBRIDGE SCHOLARS
PUBLISHING

News as Changing Texts:
Corpora, Methodologies and Analysis,
by Roberta Facchinetti, Nicholas Brownlees, Birte Bös and Udo Fries

This book first published 2012

Cambridge Scholars Publishing

12 Back Chapman Street, Newcastle upon Tyne, NE6 2XX, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-4438-3566-8, ISBN (13): 978-1-4438-3566-4

CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

ROBERTA FACCHINETTI

> *Newspapers are not a dying breed, although they are metamorphosing under pressures of technological developments and operating in increasingly competitive markets for readers and revenues. Newspapers are experiencing fundamental changes in their formats and contents, their economic organization and finance, the newsgathering and reporting practices of the journalists who deliver the news, as well as the technology employed to produce and distribute them.*
> (Franklin 2008: 30).

The mainstay of this book is the dialectic interrelation between 'news' and 'change', whereby news is intended as a textual type in its evolutionary—and revolutionary—development, while change will be discussed with reference to the form, content and structure of such typological variety explored across the centuries, largely in the British environment. Bearing in mind that any change often has a hard life, at least in its early stages, and that it often has mistrust as its twin brother, we will also illustrate the problematicity of such change(s) focusing on three main dimensions: corpora, methodology and analysis.

Indeed, when, some fifty years ago, the compilers of the first computerized corpora posited that the analysis of language in use would open up a new dimension of linguistic research, a few scholars heard their generativist foundations crack and tried to raise a Chomskyan wall against this linguistic breakthrough, which advocated a shift from the sheer paradigmatic level of linguistic analysis to the syntagmatic one. As a result, a dichotomy between corpus linguists and the so-called 'armchair linguists' took shape. Luckily, the two opposing sides have now started to steer the middle course, that is, to merge the positivity of both theoretical and corpus-based research, thus leading to great progress in the field.

Similarly, when newspapers took their first steps back in the early seventeenth century, a mechanical printing press was the only available tool. Then, technological breakthroughs started to gain ground at a

tremendous pace; the amelioration of the printing press first, the invention of the telegraph afterwards and finally the birth of the Internet were leaps of the mind into new dimensions of news making and publishing, and in the end they have contributed to ensuring global circulation of news in our present highly networked world.

However, while the telegraph was greeted with joy by the news making professionals, the Internet was first regarded with suspicion, because it meant an epochal shift into the wired environment with news visualized on screen, to be 'clicked in'—and out—at the blink of an eye. In the same way as generativists rose up against corpus-linguists, a good part of print journalism reacted against online fans. This tug-of-war is not over yet, but officially the two channels of news production have reached a truce by joining forces and now mainstream print media are even endorsing online citizen journalism.

Through the centuries, news itself, as a textual type, has not forded unharmed the rivers of ink; on the contrary, it has continuously adapted to new environments. At its birth, in the seventeenth century, only reports were published; in a matter of years, news stories were delivered with a clear angle mirroring the attitude either of the editor or of the writer (who, at first, were actually the same person); then the attitudinal slant took a life of its own and gave birth to fully-fledged commentaries and features, while news reports consolidated their top-down structure. Meanwhile, the increasing amount of information gathered in newsrooms contributed to re-shaping newspaper sections and texts; and finally, from pictures to photos, from videos to sound, from intra-textual to inter-textual links, the 'news piece' gave way to the 'news package'.

All such changes in news forms, structure and content have occurred dialectically through the centuries and are still ongoing; indeed, in the current media-saturated world, the inverted pyramid is now being challenged by more 'featuresque' intros and the neutrality of news reporting is being overshadowed by perspective and subjectivity, to the point that even the clear-cut division between reports, features and commentaries is now somehow blurred.

The present book attempts to capture all the above-mentioned changes in a diachronic perspective. Specifically, in the first chapter Nicholas Brownlees focuses on the first steps of news reporting between 1620, when newssheets and pamphlets started to be published at frequent if not totally regular intervals, and 1665, the year marking the founding of *The Oxford Gazette*, which is widely acknowledged to be England's first newspaper. In the second chapter, Udo Fries covers the following century (1665-1765), which saw an unprecedented rise in the number of news

publications all over the British Isles, while in the third chapter Birte Bös covers two full centuries (1760-1960), when political, economic and social changes, let alone demographic and cultural aspects, but also legal and institutional constraints as well as technological innovations, contributed to making the news writing landscape even more complex and diversified. The closing chapter, Chapter 4, by Roberta Facchinetti, focuses on the latest fifty years of news reporting, starting from the 1960s, a decade that most marked technological enterprise in the 20th century, up until the present time, that sees the monovocality of mainstream newsreporting challenged by the multivocality of social networks.

The uneven distribution of the time spans in the four chapters has been decided on account of the following two concomitant factors: (a) historical key moments in the process of news writing changes, particularly with reference to Chapters 1 and 4; (b) extant computerized corpora covering such periods, thereby permitting specific linguistic analyses, particularly for Chapters 2 and 3, respectively focusing on the ZEN corpus and on the Rostock Newspaper Corpus.

Indeed, to investigate the structural and linguistic peculiarities of news reporting through the centuries, in each chapter we make use of a set of corpora specifically devised to suit the needs of scholars studying the periods under scrutiny. In Chapter 1, the focus is particularly on the representation—in early periodical news—of (a) authorial presence and identity, mostly the function of personal pronouns and of (b) heads and marginalia. The analysis of Chapter 2 is dedicated to the diversification of textual types also with reference to lexical and stylistic comparisons, use of foreign words and phrases, headlines and text beginnings, speech and thought presentation and also graphemics. Chapter 3 discusses the development of popular journalism, between New Journalism and Tabloidization, and illustrates attempts to (a) classify newspapers on a popular/quality-scale and (b) analyze linguistic aspects leading to the 'conversationalization' process, mostly first and second person pronouns, the use of the imperative and of quotations. Finally, in Chapter 4, attention is given to the mutable/mutating linguistic specificities of news reporting largely on account of (a) the influence of the language of social networks and unmediated journalism, (b) the new structure of news packages and (c) the increasing swing between 'impartiality' and 'perspective' in the news output.

The topics discussed and the corpora exploited to analyze them call into question basic methodological issues that are tackled from different perspectives in each chapter. Indeed, on the one hand, corpus-assisted discourse studies (CADs) lead us to inductively investigate and test any

research question against original data. On the other hand, we are also aware that corpus design poses different problems depending on the research aims and on the interpretation of results. Hence, we are particularly keen on the fact that whatever corpus is exploited, both quantitative and qualitative analyses have to be taken into account, quantitative data being the starting point for further qualitative analysis and, in turn, qualitative analysis being the spur for further quantitative investigation. The epicentre of such a never-ending circle in research is the news itself, while researchers profit from cooperation among themselves.

This tenet has definitely proved right for the co-authors of this book; indeed, each chapter has been written by each one of us according to our professional competences, but has also been read and revised by all co-authors, to ensure consistency and mutual enrichment. The experience of writing this book together has certainly enriched the four of us and we hope it will enrich any scholar, student and professional or non-professional journalist who comes across the fascinating world of news in their everyday life.

CHAPTER ONE

THE BEGINNINGS OF PERIODICAL NEWS (1620-1665)

NICHOLAS BROWNLEES

1. History of periodical news (1620-1665)

My intention in this chapter is to examine specific linguistic and structural features of early English periodical news. In particular, I shall consider, first, the representation of authorial presence and identity and, secondly, the role of heads and marginalia in the presentation of periodical news. Both these research questions are analyzed with the aid of two computerized news corpora that will be described in Section 2 of the present Chapter.

As regards the time span covered in this analysis, I will be focusing on the period between 1620-1665. In these years both the initial and final date are crucial in the history of the English press, since, whilst the former represents the moment when news publishers set about publishing first newssheets and then pamphlets at frequent if not totally regular intervals, the latter year marks the founding of *The Oxford Gazette*, the twice weekly publication that in format, content and periodicity can be considered England's first newspaper.

A study of these first four decades of English periodical news is as fascinating as the momentous times that were being reported in the various newssheets and weekly pamphlets themselves. What we see is a news discourse grappling with the one fundamental question lying at the heart of all news communication: how should news be presented? What linguistic, interpersonal and graphic features should be adopted to convince the reader to buy the news publication from one week to the next?

In its novelty this question was as taxing for contemporary news writers and publishers then as is the role of the internet in news publication nowadays. Whilst in these first years of the twenty-first century news

professionals have to determine how far traditional practices of news transmission and presentation are valid in the era of the internet, in the first decades of the periodical press news writers had to decide how much of previous forms of news transmission could be successfully exploited in the reporting of periodical news. Prior to 1620, news had been published in the form of occasional news pamphlets, but these publications made no claim to periodicity and generally just focused on one or two relatively recent news events. Was the news idiom found in these occasional news pamphlets—that often spoke of natural disaster, crime and bloody murder, miracles and extraordinary events, pageantry and political spectacle, and more often than not within a highly moralistic framework[1]—inherently suitable for news publications that determined to come out more or less regularly, providing each time they did so an update on contemporary events and affairs? We shall see that there was no fixed answer. For the first forty-five years of its existence, that is, between 1620-1665, English periodical news presented various different discourses and styles in its attempt to find and keep an audience who would pay one or two pennies a time for the latest news of the day.

The earliest extant periodical news publication was published on 2 December 1620. Although its first words were undramatic in the extreme—"The new tydings out of Italie are not yet com"—the news publication containing this news is very significant.[2] For posterity it is the first of hundreds of publications that are most frequently referred to as 'corantos'.[3] For the first year of their existence English corantos were first printed in Amsterdam and other towns in Holland and then shipped across the English Channel, after which they were sold at various booksellers and stationers in the St Paul's area of London. These early corantos were more or less literal translations of previously published Dutch newssheets. As such, most of the news was focused on the early stages of the Thirty Years

[1] See, for example, *Newes from France. Or a relation of a maruellous and fearfull accident of a disaster* (1618), *Newes from Spain. A true relation of the lamentable accidents* (1618), *A true relation of a most desperate murder* (1617), *A wonder woorth the reading, or a true and faithfull relation of a woman* (1617), *A true relation, of the happy peace concluded by the two mighty princes, Christian the Fourth, King of Denmarke, and Norway* (1613), *A true relation of Go[ds] wonderfull mercies* (1605). The language of occasional news pamphlets between 1600-1620 is examined in Brownlees (2011: 1-24).

[2] Since this newssheet contains no title, it is catalogued under these opening words. See Dahl (1953) for a bibliography of almost all extant corantos between 1620-1642.

[3] They were variously referred to by contemporaries as 'coranto', 'curranto', 'corante', 'corant', or 'courant'.

War fanning out across continental Europe. Each of the corantos consisted of one small folio sheet, with news closely printed in two columns on both front and back page, the various news items being separated by datelines.

> Out of Weenen, the 6 November.
> The French Ambassadour hath caused the Earle of Dampier to be buried stately at Presburg. In the meane while hath Bethlem Gabor cited all the Hungerish States, to com together at Presburg the 5. of this present, to discourse aboute the Crowning & other causes concerning the same Kingdom.
> The Hungarians continue with roveing against these Lands. In like manner those of Moravia, which are fallen uppon the Cosackes yester night by Hosleyn, set them on fire, and slaine many dead, the rest will revenge the same. […]
> *Out of Prage, the 5 of November.*
> Three dayes agone are passed by, *2.* mile from this Cittie 6000. Hungarians (chosen out Soldiers) under the General Rediserens, which are gon to our Head-camp, & the Enimie lieth yet near unto ours by Rackonits, though the crie goeth, that the enimie caused all his might to com togither, to com this wayes against Prage, if that comes to passe, it shall not run of without blowes, the which might be revealed with in few dayes.
> (*The new tydings out of Italie are not yet com*, 2 December 1620)

The style of these news dispatches is that of factual, unadorned, news reportage. It is hard news with the focus on *who*, *where*, *what* and *when*. Little importance is given to *why* something happened since the various writers of the dispatches did not regard it as their duty to interpret events. They were the purveyors of news; comment and analysis were generally left to the reader.

These first corantos must have been successful because by the summer of 1621 London publishers had decided to print and sell translated Dutch and German newssheets themselves. The seven extant London-published corantos of the summer and early autumn of 1621 were published by "N.B.", initials that must have stood for Nathaniel Butter or Nicholas Bourne, two names that were to remain associated with the world of periodical news publications over the next twenty years.

In 1622 the manner of recounting periodical news changed. Whilst in the first two years periodical newssheets had principally consisted of a succession of unrelated news dispatches from diverse parts of Europe, in 1622 we see the editor intervening in an attempt to provide a more coherent account of the latest news. One of the first occasions we see evidence of this is in June 1622. In the passage below the editor explains a new approach to the writing up of news.

> Wee write a continuation, that you may see by the proceedings, that there is good dependancy betweene the relations, wherein we purpose to keepe nere to the Lawes of Historie, to guesse at the reasons of the actions by the most apparant presumptions […]
> (*A continuation of more newes from the Palatinate*, 13 June 1622)

The editor *cum* news writer (for they were generally the same person) is proposing to guide the reader from one publication to another through the news stories of the day. In so doing, the editor frequently intersperses his factual recount of the news with more general comment upon what the news signifies in a wider framework of human action and behaviour:[4]

> The whole countrey thereabouts, with the Bishopricke of *Spyres*, was left to the deuotion of the King of *Bohemiah*, now by this victory of *Hagenaw*, made sole master of the field; whereupon the *Mansfeldians* made what rauage they pleased in that delicate and most fruitfull countrey: And the Count *van Hannow* of the lower *Alsatia* (neighbor to *Leopoldus*) taking into due consideration, that rule of warre; When two powerfull Princes are either declared or engaged in a warre, it hath beene still obserued to be a dangerous thing, for any third man to stand as a neutrall, who is not able against either of them to stand of himselfe: For still the Victor when he hath ouerthrowne the enemies, falls next vpon the weaker Neuters: So that if the Victor has leisure afterward to prosecute them, it is as safe almost, for to haue beene a dependant on the vanquished, as to haue beene onely a looker on
> (*A continuation of more newes from the Palatinate*, 13 June 1622)

Quite often, we find the editor using domesticating strategies designed to facilitate the understanding of far off events in remote European locations. Two such strategies involve the use of metaphor and proverb.

> To begin at the head, whence the beginnings of all motion, influence, and direction to the whole body, and the severall parts and organs proceed; We thinke it manners first to relate the businesse of the Emperiall Maiesty, vpon whose deliberations & resolutions, the actions of those parts of the world haue their dependancy
> (*A relation of the last newes from severall parts of the world*, 8 April 1623)
> These sixteene are to be executed as they can be taken. And if the Prouerbe (as I thinke) tooke his beginning from this custome, That hee that hath lost

[4] As the world of early seventeenth-century English news writing was peopled by the male sex, my use of the third-person male pronoun extends generically to news writers as a whole at that time.

> his good name is halfe hanged; then are these Gentlemen whose names are thus ignominiously fixt vpon the gibbets, sure to suffer for it
> (*A relation of the weekely occurrences*, 22 October 1622)

As this popularizing narration of news is evidence of editorial intervention, it is no wonder that the editor concerned was well known to the news reading public of the time. His name was Thomas Gainsford, and between the summer of 1622 and his death from the plague in the late summer of 1624 he frequently wrote up the news of the time in a highly personal voice.

After his death, the news idiom once again reverted to the dry, matter-of-fact dispatch style of foreign newssheets as once again the two main coranto publishers, Nicholas Bourne and Nathaniel Butter, relied heavily on the simple translation of foreign news reports for their own English publications. Whether this decision was determined by financial considerations, or by a belief that this impersonal news style was preferred by the reading public, or by the realization that such unmediated news was least likely to offend the ever watchful government authorities of the day, is uncertain, but whatever the case most periodical news was characterized by an impersonal, heavily factual mode of narration up until the end of 1641.[5]

What led to the change in news discourse was the fast moving political situation overtaking British political life in the first years of the 1640s. Central rule and governance broke down as the king, Charles I (1625-1649), found himself unable to impose his authority upon a parliament ever more determined to question the inherent legitimacy of royal prerogative. In this battle of wills, and later arms, one of the first features of royal power to disappear was that of censorship. In 1641 the various forms of censorship exercised by Star Chamber, the crown's much feared and hated court of law, were abrogated as Star Chamber itself was abolished.[6] The result was that for the first time English periodical news publications were no longer limited to the publication of foreign news unconnected to British affairs.[7] However, not only did English news

[5] See Brownlees (2011: 25-96) for an analysis of discourse features in corantos between 1620-1641.

[6] See McElligott (2005) for how censorship affected the publication of news in early modern England.

[7] The only corantos to provide detailed news affecting English interests abroad were published in the summer of 1627 at the time of the duke of Buckingham's ill-judged invasion of France. See, for example, *The continuation of our weekely newes* (1 August 1627).

pamphlets, or 'newsbooks' as historians frequently refer to these periodical publications between 1641-1665, start publishing national news but by 1643 they also began adopting a very different kind of news discourse from what had been adopted in the first decades of periodical news.[8] This new mode of news language is seen in the following passage, taken from *Mercurius Aulicus*, a newsbook that began publication in January 1643.

> This evening by an expresse we received advertisement that the Rebels Army approached *Banbury*, and gave an Alarum to the Towne; whereupon the Earle of *Northamptons* two noble Brothers and Lieutenant Colonel *Greene* (who command the Castle and Garrison in the Earles absence) hosted 80 men, who went out as farre as *Adderbury*, where the worthy Lord Gray was enquartered, and there seized upon 18 of the Rebels Souldiers, which they brought into the Garrison at *Banbury*, to learne better obedience.
>
> All other Newes (I mean Lyes) you must expect from a fine new thing, borne this weeke, called *Mercurius Britannicus*, for Mercuries (like *Committees*) will beget one another. But sure he is no true Brittaine, for the first thing he said, is, that *most of the* Welchmen *for piety and godlinesse are as ignorant as Heathens* (the man begins handsomely. (2. He sayes, that *the* Irish *Rebels doe daily land in* Wales, *and are there made welcome*: (They'le doe as much for you, if you'le goe thither.) 3. That *the Earle of* Ormond *by his delayes hath put many thousand Protestants to sword*, (and yet you say the *Rebels hate him as their most active enemy*.)
> (*Mercurius Aulicus*, 27 August-2 September 1643)

What we see in this passage are two distinct types of news discourse. The first kind is found in the first paragraph, where the news writer is reporting recent events in the Civil War that had broken out between the king and parliament in August 1642. Although the news content is domestic, and hence in contrast to the foreign news dispatches making up coranto news in the first two decades of periodical news, the style of narration is not so dissimilar from the earlier mode of news narration. Subjective involvement in the text is limited to the exclusive personal pronoun "we" in line 1, no attempt being made to construct an interpersonal relationship with the reader. In contrast, in the second paragraph, not only do we see the invective and polemic that came to typify much Civil War and Interregnum news discourse but also both subjective (*I*, *you*) and intersubjective (*he*) personal pronouns where *I* refers to the news writer, *you* to both the reader ("you must expect from a fine new thing") and the

[8] For monographs containing extensive discussion on 1640s and 1650s newsbooks, see Frank (1961), Raymond (1996; 2003) and Peacey (2004).

cited newsbook ("and yet you say the Rebels hate him") and *he* to the personalized adversarial newsbook *Mercurius Britanicus*).[9] These different markers of authorial involvement and intertextual communication characterize a discourse strategy that became the hallmark of many news texts during the Civil War years. Called *ad locum* animadversion in classical rhetoric,[10] and 'disclaiming through engagement' by Martin and White (2005: 118-121), the discourse strategy involves the refutation of a previously produced work by means of first citation and then rebuttal of the prior text. This form of dialogic confrontation was a common feature of the 'Pamphlet Wars' or 'War of Words' that took place in the 1640s, as news writers of diverse political and religious leanings exploited this unique moment in seventeenth-century English history, when censorship was for practical purposes non-existent, to question, criticize and frequently deride other periodical news texts crowding the London booksellers' shelves.

The heady explosion of press freedom came to a halt in the autumn of 1649. It was then that Oliver Cromwell, by now *de facto* leader of the victorious parliamentarians (Charles I having been executed in January 1649) passed a law that drastically reined in the number of periodical news publications. While initially the only newsbooks to survive for any length of time were government approved, by October 1655 the newsbooks were not just authorized but directly written on behalf of the government. The man who was granted sole permission to print periodical news for the Protectorate, as the English government was then known, was Marchamont Nedham, a colourful character and highly gifted polemicist and editorialist. Formerly editor of the highly politicized 1640s newsbooks *Mercurius Britanicus* and *Mercurius Pragmaticus*, he also displayed the more traditional virtues of good journalistic attention to detail and accuracy in his 1650s newsbooks *Mercurius Politicus* and *The Publick Intelligencer*. Making good use of the information regularly supplied to him by his excellent contacts at home, and his creation of a highly efficient network of correspondents abroad, he generally provided good quality news.[11]

---

[9] Although spelt 'Mercurius Britannicus' in the cited passage from *Mercurius Aulicus*, the true title of the parliamentarian newsbook was *Mercurius Britanicus*, hence, without the second 'n' as is the case in the correct Latin spelling. *Mercurius Britanicus* was misspelt in the first number, and as such the title remained.

[10] See Raymond (2003: 211) for a typology of the various forms of animadversion frequently employed in Civil War pamphlet polemic.

[11] See Frank (1980) for a critical biography of Marchamont Nedham.

Regarding the content of home news, what distinguishes the 1650s from previous decades is an increased use of advertisements and a greater focus on crime and court proceedings.[12] The first two passages below typify the ever growing book and medical advertisements found in periodical publications at that time whilst the third passage is an example of a crime report.

> ☞ *The Good Old Way*; or Perkins *Improved*:
> In a Plain Exposition and sound Application of those depths of Divinity briefly comprised in his six Principles: by that late painful and faithful Minister of the Gospel, *Charls Broxolm*, in Derbyshire. To be sold by *John Rothwell* at the Fountain and Bear in *Cheapside*; and by *Joseph Barber* at the Lamb in *Pauls* Churchyard.
> (*Mercurius Politicus*, 3-10 November 1653)

> People that are Melancholy or Distracted, are kept and preserved from danger in a very convenient place for that purpose, being in an excellent Air highly commended, likewise for those that are in Consumption. Such whose Cure is intended, may have the advice of Physitians usually frequenting the House. They that have occasion may inquire farther of Mistress *Jackson* in Drury-Lane, near the *Fortune* Tavern over against *Long Aker*, or of Mr. *Field* an apothecary near the Gatehouse of *Westminster*.
> (*Mercurius Politicus*, 16-23 October 1656)

> The said *Jane* being walking in *Greenwich* Park, with Mrs. Smith her Kinswoman, and two maids attending her, the said Master *Welsh, Robert Thompson*, and others, armed with Swords, Pistols, and other weapons, in pursuance of that design did enter the said Park, and being all strangers to the said *Jane*, violently seized upon her; and forcibly set her on horseback, one also violently threw Mistris *Smith* upon the ground, set his knee upon her breast and kept her down, while the others laid hold of her two maid servants, and withheld them from helping of her. [...]
>
> The Charge being read, the Court ordered that the said Mr. *Welch*, should personally appear on Saturday come sevennight being the 15 of this instant March at Sergeants Inne aforesaid, at 9 of the clock in the forenoon, to make his defence to the said Charge and pretended marriage by him, and

[12] The earliest advertisement in a coranto appears on 16 September 1624 (Dahl 1953: 125). For advertising in newsbooks, see Frank (1961: 201-202, 256-257 and *passim*) and Sommerville (1996: 54-55). See Chapter 2 of the present volume for extensive analysis of the role of advertising in periodical news publications between 1665-1765.

> have likewise ordered summons to be sent forth against Mrs. *Horwood*, Mrs. *Basset*, and others, to appear at the same time.
> (*A Perfect Account of the Daily Intelligence*, 5-12 March 1651)

When Charles II was restored to the throne in 1660, it was no surprise that Marchamont Nedham was not around to report the popular celebrations. He had escaped to Holland, and in his place the royalist Henry Muddiman had been installed as official publisher of periodical news. However, whilst Muddiman and following him Roger L'Estrange now took over what had been previously carried out by Nedham, what did not change was the essential news product that was being produced. It was still a newsbook consisting of between 8 and 16 quarto pages. It was only in November 1665, with the first number of the *Oxford Gazette*, that the format changed, thereby heralding the end of periodical pamphlet news.

2. Corpora

The two machine-readable corpora I shall examine for this period are the *Florence Early English Newspapers* (FEEN) *Corpus* and the *Lancaster Newsbooks Corpus*. The FEEN corpus covers the period from December 1620 until April 1653 whilst the *Lancaster Newsbooks Corpus* spans the period from the middle of December 1653 until the end of May 1654. At present I am unaware of any electronic corpora for the period from 1654-1665, the second year being when the *Oxford Gazette* began publication. However, whilst it would be statistically useful to have further electronically readable texts for this period, it is unlikely that their examination would offer any important new insights which cannot already be gleaned from the 1650 corpora available. The next major change in English news production and language occurred with the introduction of the *Oxford Gazette* in 1665 (renamed *The London Gazette* in 1666), and this publication is included in the *Zurich English Newspapers Corpus*, which is examined in Chapter 2 of this present volume.

The FEEN corpus is divided into six subcorpora designed to represent particular aspects of periodical news publication between 1620-1653. The first group consists of the corantos published between December 1620 and October 1621. These are more or less literal translations of Dutch or German newssheets previously published in continental Europe. The second subcorpus is made up of corantos published between the summers of 1622 and 1624. As has been stated above, these two years were especially significant in the early history of English periodical news discourse. It was then that a marked effort was made to forge a new discourse style, a model of news presentation that was very different from

the generally laconic news dispatches that were found in Europe and which had made up the first English corantos. The final group of corantos includes news publications published between 1625 and 1641. Bearing in mind that between October 1632 and the end of 1638 corantos were banned in England, this means in practice that the corantos range from 1625-1632 and 1639-1641.

The fourth and fifth subcorpora consist of 1640s newsbooks. The smaller of these subcorpora include several well-known newsbooks of the decade: *Mercurius Civicus*, *Mercurius Pragmaticus*, *The Kingdomes Weekly Intelligencer*, *The Moderate* and *The Moderate Intelligencer*. The larger 1640s subcorpus is, instead, made up of 22 numbers of *Mercurius Aulicus* and 16 numbers of Marchamont Nedham's *Mercurius Britanicus*. *Mercurius Aulicus* (henceforth *Aulicus*) was the preeminent royalist newsbook, the "Aulicus" in the title referring to the court of the English monarch, Charles I. Published in Oxford, though there were also occasional London reprints, like other newsbook series it came out once a week. In contrast to *Aulicus*, the London-published *Mercurius Britanicus* was parliamentarian in outlook. It was founded in August 1643, eight months after *Aulicus*, to counter the latter's highly effective propaganda. Indeed, the presence of "Britanicus" in the title of the parliamentarian newsbook underlines its intention to report and defend the news and rights of the nation at large. The decision to focus on *Aulicus* and *Mercurius Britanicus* (henceforth *Britanicus*) in this subcorpus was not just motivated by the relative importance of the two news pamphlets but also by the extent to which the two news publications directly interact with each other. Ten of the *Aulicus* numbers in the FEEN corpus have specific sections devoted to attacking the news of parliamentarian pamphlets, and in particular the news published by *Britanicus*. This same structure is reflected in *Britanicus*, which is also often divided into two parts, with the first section consisting of an account of the week's news and the second a commentary of the enemy's news. The latter section is entitled "Aulicus" in that Nedham exclusively attacks the Oxford pamphlet. This subcorpus, therefore, provides a useful sample of Martin and White's above-mentioned 'disclaiming through engagement' strategy.

The final subcorpus of FEEN consists of three titles, and seven publications, published between 1650-1653. The newsbooks comprising the subcorpus are *Mercurius Politicus*, *The Perfect Diurnall of Some Passages and Proceedings* and *A Perfect Account of the Daily Intelligence*. In these years the first two titles were the most important but *A Perfect Account of the Daily Intelligence* is also included in the corpus since it is described as "extremely typical" of the times (Frank 1961: 241). In

conclusion, the dataset in the FEEN corpus were not selected at random (as is sometimes quite justifiably the procedure with some corpora compilation) but were rather the result of an informed acquaintanceship with early English news texts in general.

In the FEEN corpus the original corantos and newsbooks are found at the British Library in the *Burney Collection of Newspapers* and *Thomason Tracts*. The corpus texts had to be keyed in manually because the quality of the original images was too low to permit satisfactory OCR scanning. Regarding manual digitization, Baker (2006: 35) writes it is "the final, and usually last resort of the corpus builder", and he is right, since it is either extremely labour intensive or costly (should you pay someone to do it professionally), but if the keying in is carried out by the compiler it does at least give him or her the chance to become not just acquainted with the text but also aware of the kinds of research questions that could prove worthy of investigation.[13] In FEEN the corpus containing *Aulicus* and *Britanicus*, the fifth and largest subcorpus, was compiled by Francesca Benucci and Nicholas Brownlees, whilst the other five subcorpora were compiled by Brownlees. Excepting those cases where the original text is unclear, all the single news texts in FEEN are complete copies of the original publications. The compilers decided to focus on a limited range of complete numbers rather than a much greater range of samples of individual publications on the grounds that at various times between 1620-1641, and increasingly in the 1640s, news texts became increasingly heterogeneous in content and style (Raymond 2003: 214).

At present the subcorpus which has the most developed electronic formatting is that containing the *Aulicus* and *Britanicus* publications. This is marked up in XML (Extensible Markup Language), a markup language based on the Standard Generalized Markup Language (SGML) which is adopted for encoding electronic texts. Each XML file included in the *Aulicus/Britanicus* subcorpus has a corresponding DTD (Document Type Definition) file defining its markup characteristics. These include the title, week number, date line, page number, and, in the case of *Britanicus*, the margin captions that are frequently found. The other five subcorpora in FEEN are in the process of being marked up in similar form. It is recognized that this level of annotation is quite basic but for the time being it is considered sufficient for the kinds of research questions being posed. Furthermore, as Baker says (2006: 42), it is always "possible to go back to the building stage at a later point to carry out new forms of annotation".

[13] The advantage of knowing one's corpus is underlined by Partington (2003: 12) who ensured that his corpus of downloaded White House press briefings was not so large that he could not read it in full.

Indeed, in the compilation of a corpus, I agree with Rissanen (1989: 17) who believes that a corpus should be kept "open-ended". It should be so structured that improvement and supplementation are straightforward operations.

At the moment FEEN can only be accessed privately, but once the markup is complete the intention is to place the corpus in the public domain. This will not only give other researchers the possibility of analyzing different features in the corpus from what are examined in the present study but also of checking the present findings. This is important from a methodological point of view since a researcher's findings are only truly scientifically valid if they can be replicated and corroborated by independent observers (Stubbs 2001: 123).[14]

Period	Contents	No. of publications	Word count for period
1620-1621	Corantos translated from Dutch/German newssheets	4	8,000
1622-1624	Corantos with much editorial input	5	21,500
1625-1641	Corantos generally based on foreign dispatches or translations	9	28,300
1642-1649	7 separate titles apart from *Aulicus* and *Britanicus*	12	40,000
1643-1644	*Aulicus* and *Britanicus* newsbooks	38	148,400
1650-1653	3 separate titles	7	32,600
Total		**75**	**278,800**

Table 1.1: FEEN (Florence Early English Newspapers) corpus. Figures in Table 1.1 and Table 1.2 are rounded up to hundreds. See Appendix for titles of publications in FEEN.

[14] See, however, Partington (2009: 293-294) who argues that rather than relying exclusively on 'pure replicability' it is often more practical to base scientific validity on the concept of 'para-replication'.

The *Lancaster Newsbooks Corpus*, which can be accessed free of charge at *The Oxford Text Archive*, is very different in size, focus and mark up from the FEEN corpus.[15] First of all, it is significantly larger in that it amounts to one million words. Secondly, rather than including selected texts over an extended period of time it includes a continuous series of news publications over a short period of time. Thus, it not only comprises the full run of the satirical newsbook *Mercurius Fumigosus* (1654-1655), but the complete collection of every mainstream newsbook published in London between the middle of December 1653 and the end of May 1654. This latter part of the corpus amounts in all to 870,000 words and consists of 23 different titles (Prentice and Hardie 2009: 31). However, since the full corpus was considered unnecessary for the purposes of my research questions in this present study, I limited my examination of the *Lancaster Newsbooks Corpus* (henceforth called LNC) to three newsbooks. They are respectively *Mercurius Politicus*, *The Perfect Diurnall of Some Passages and Proceedings* and *A Perfect Account of the Daily Intelligence*—the same three newsbooks already found in the 1650s subcorpus of FEEN. The word count for these newsbooks that I have selected from LNC, coupled with the 1650s newsbooks in FEEN, is approximately the same as that for the 1640s FEEN newsbook corpus.

The data encoding system employed in LNC is explained at the project web site.[16] Apart from information as to the inputting format, which is described as a "basic SGML-compatible format that has much in common with HTML", we are also given details regarding principal markup procedures. They include text elements signalling extent of paragraph, the three main typeface variants (roman, italics, gothic), text types such as 'table' and 'poem', page breaks, and where there is unclear text. What is also interesting is the decision to provide both the original spelling and a standardized modern-day English orthographic form in the markup. LNC's project leaders, Tony McEnery and Andrew Hardie, explain that this facilitates concordancing since through a simple algorithm it is possible to identify alternate spellings of words, a common phenomenon in mid-seventeenth century England. Hence, by means of this markup procedure, it is possible to recognize that *Cromwell* and *Cromwel* (an alternate spelling of the name) are in fact the same word. Although not stated in LNC's explanation, this markup facility above all helps researchers with little experience of spelling practice of that time. For those with greater acquaintanceship of seventeenth-century news texts, and texts of that time

[15] A comprehensive list of online archives and corpora (including links) is provided at the end of this volume.
[16] http://www.lancs.ac.uk/fass/projects/newsbooks (accessed April 2011).

in general, the resource, while still useful, is not essential since the researcher will be already aware of the likelihood of possible alternate spellings, and can thus make full use of concordancing software.

Period	Contents	No. of publications	Word count for period
1653-1654	*Mercurius Politicus* (newsbook)	23	118,700
1653-1654	*A Perfect Account of the Daily Intelligence* (newsbook)	12	31,200
1653-1654	*The Perfect Diurnall of Some Passages and Proceedings* (newsbook)	3	16,500
Total word count		**38**	**166,400**

Table 1.2: Dataset for present study taken from *Lancaster Newsbooks Corpus*.

Concluding the description of the corpora accessed in the present study, we see that the total word count of FEEN combined with the dataset selected from LNC therefore comes to 445,200. This figure comprises 199,000 words of 1650s newsbooks, 188,400 of 1640s newsbooks, and 57,800 words of corantos between 1620-1641. The 1650s and 1640s word counts are not insubstantial and certainly give the researcher the possibility of identifying linguistic features and semantic collocations in some of the main news publications of the period. The corantos' word count is clearly much smaller but one needs to remember that not only were there no such publications between the autumn of 1632 and December 1638 but many fewer corantos were published than newsbooks. Of the 404 corantos recorded in Dahl's coranto bibliography (1953), no more than 350 are extant, which means that although small in number the corantos in FEEN still amount to more than 5% of the total extant output. Furthermore, the representativeness of the FEEN corantos is enhanced by the subdivision of the texts into three subcorpora. Through these subcorpora we have the opportunity of recognizing specific discourse styles particular to certain periods over the two decades.