

Subordination as a potential marker of complexity in serious and popular fiction: a corpus stylistic approach to the testing of literary critical claims

Rocío Montoro:

University of Granada
Facultad de Filosofía y Letras.
Departamento de Filologías Inglesa y Alemana
Campus de Cartuja, s/n.
18071 Granada (Spain)
rmontoro@ugr.es

Dan McIntyre:

Department of Linguistics and Modern Languages
University of Huddersfield
Queensgate
Huddersfield HD1 3DH (UK)
d.mcintyre@hud.ac.uk

Abstract

In this article we use a corpus stylistic methodology to investigate whether serious (i.e. 'literary') fiction is syntactically more complex than popular (i.e. 'genre') fiction. This is on the basis of literary critical claims that the structural complexity of serious fiction is one of the features that distinguishes it from popular literature (which, by contrast, is seen as easier to read). We compare the serious and popular fiction sections of the Lancaster Speech, Writing and Thought Presentation corpus (see Semino and Short, 2004) against various samples of the British National Corpus available in Wmatrix (Rayson, 2009), focusing particularly (though not exclusively) on the identification of subordinating conjunctions. We find that, on this measure, there is no basis for claiming that serious fiction is any more complex syntactically than popular fiction. We then investigate the issue in relation to a specific genre of popular fiction, Chick Lit. Here we find that while syntactic simplicity exists, this is at a phrasal rather than clausal level. We argue that by using a corpus stylistic approach we are able to qualify accurately certain literary critical claims about syntactic complexity as a distinguishing feature of serious and popular fiction, and to propose a refined hypothesis which might be used in further studies of the syntactic structures used in these two text types.

Keywords corpus stylistics, Chick Lit, serious and popular fiction, subordinating conjunctions.

Subordination as a potential marker of complexity in serious and popular fiction: a corpus stylistic approach to the testing of literary critical claims

Rocío Montoro and Dan McIntyre

1. Introduction

This article uses corpus stylistics to analyse serious (also referred to as 'literary' or 'high-brow') and popular (often called 'genre' or 'low-brow'¹) fiction in relation to one claim often made about the latter, namely its apparent simplicity when compared to the former. As discussed below, literary criticism often highlights the apparent structural complexity of serious fiction as a feature that distinguishes it from popular literature which tends to be seen as easier to read. For instance, on the style of David Foster Wallace, de Bourcier (2017: 1) comments: 'The length and complexity of Wallace's sentences has often been remarked on'. From a stylistic perspective, prose fiction was easily the most studied genre of the 20th century but most of this research was on the stylistics of serious fiction. In contrast, the language of popular literature has been relatively under-researched. Recent work in stylistics (e.g. Stockwell, 2000; Mandala, 2010; McIntyre, 2012; Montoro, 2012) has begun to build on early studies by scholars such as Nash (1990), although there is still a considerable amount of work to do to determine whether the commonly held distinctions between serious and popular fiction are reflected linguistically. As a contribution to this ongoing research agenda, we investigate the language of popular literature from a primarily corpus stylistic perspective and consider whether and, if so, how the presumed complexity of serious fiction is linguistically realised. In this article we focus exclusively on syntactic complexity as the notion of 'complexity' is in itself a much broader issue which, due to space constraints, we cannot develop fully here². Specifically, we focus on syntactic complexity which is marked by subordinating conjunctions although, of course, we are not claiming that syntactic complexity is projected exclusively by subordination, nor subordination realised simply by this means (Montoro (2018) develops the issues addressed in this paper by considering claims of syntactic simplicity in popular fiction with regard to nominal post-modification). Our analysis is rather an attempt to account for the popular vs. serious fiction distinction in a more rigorous way than has been so far undertaken. Thus, the hypothesis informing this paper is that complex sentences will be more prevalent in serious fiction than in popular fiction. This hypothesis arises from commonly held assumptions that writers of popular fiction are less linguistically versatile than literary authors and that this is likely to be reflected in the syntactic structures they choose³.

One further remark that needs to be made is that we look at two text varieties (serious and popular fiction) which, by no means, represent coherent and discretely defined categories. That is, the issue of what constitutes literary or genre fiction is, in

¹ We are aware that these terms are not fully interchangeable and that the distinction high/low fiction, especially, might carry particular evaluative connotations. In this paper, however, we use them as approximate synonyms.

² For instance, the notion of complexity has also been addressed from a thematic, rhetorical or narrative perspective (see Gelder [2004] below). We are concerned exclusively with syntactic complexity and, specifically, with complexity realised by subordinating conjunctions.

³ See, for example, Pullum's (2004) criticisms of Dan Brown's writing style, and evidence from Montoro (2012) who finds that readers tend to associate low-brow fictional narratives with an apparent structural simplicity.

itself, also debatable and open-ended⁴. In order to minimise the danger of circularity, we use a corpus of texts which were categorised as either 'serious' or 'popular' by Semino and Short (2004). Our corpus is a sub-set of the Lancaster Speech, Writing and Thought Presentation corpus (SW&TP) (Semino and Short, 2004; available from the *Oxford Text Archive* [OTA]: <http://ota.ox.ac.uk/>), specifically the fiction section of the corpus which contains approximately 40,000 words of serious fiction and an equal amount of popular fiction. To isolate complex sentences, we tagged the corpus for part-of-speech (POS) information using the CLAWS tagger in Wmatrix (Rayson, 2009). This allowed us to search for the subordinating conjunctions introducing complex sentences, although we are aware that subordination does not work only at this level nor solely by these means. We then compared the two sub-corpora against several reference corpora, as described below. This initially allows us to determine whether or not subordinating conjunctions are (a) statistically key in the serious fiction sub-corpus (that is, over-represented in comparison with the distribution of subordinating conjunctions in the reference corpus), and/or (b) statistically under-represented in the popular fiction sub-corpus. From the results of this statistical analysis we draw preliminary conclusions concerning whether it is possible to validate or invalidate the hypothesis stated above. Following this initial analysis, we follow up with an investigation of a larger corpus of approximately 500,000 words from a very specific sub-genre of popular fiction, namely Chick Lit (Ferriss and Young, 2006). From this, we aim to determine, albeit tentatively, whether the findings obtained from the small corpus analysis are reflected in a larger sample of genre-specific writing. We discuss whether or not our initial hypothesis is supported, we provide the results of the larger corpus analysis of Chick Lit, and offer a qualitative analysis of sentence types from both serious and popular fiction in order to demonstrate the choices and effects of particular syntactic structures.

2. Popular vs. serious fiction

Debates on the notion of 'literariness' are not new in literary criticism; in fact, some authors (e.g. Burke, 2014) claim that they go as far back as discussions on poetics and rhetoric in ancient Greece. The difference between what are traditionally considered serious and popular genres has also been widely discussed and these discussions have not necessarily always led to generalised agreement. McCracken (1998), for instance, describes fiction written under the general, overarching, 'popular' denomination as, 'fiction that is read by large numbers of people' (McCracken, 1998: 1). According to McCracken, popular fiction may initially be defined by sheer size of readership; that is, popular novels are, for the most part, characterised because they tend to attract a much larger readership than serious fiction. For instance, in a list of the top 100 fiction authors for the year 2010 (Nielsen Bookscan⁵), we find novelists such as Stieg Larsson, James Patterson, Lee Child, Jodi Picoult, Dan Brown, Danielle Steel, Martina Cole, Sophie Kinsella, Harlan Coben, Philippa Gregory, Charlaine Harris, David Nicholls, Maeve Binchy, Bernard Cornwell, Sebastian Faulks, Michael Connelly, Jo Nesbo, Kathryn Stockett, Patricia Cornwell, and Karen Rose, to mention but a few. The list encompasses authors who could be categorised under such sub-genres as, for instance, crime fiction (Stieg Larsson, James Patterson, Patricia Cornwell), thriller (Lee Child, Jo Nesbo), mystery and suspense (Karen Rose, Harlan Coben), Chick Lit (Sophie

⁴ Werner (2018: 8) argues, though, that the boundaries separating what is considered 'high' and 'low' culture are being increasingly dissolved.

⁵ Nielsen Bookscan kindly provided this information for us. We thank David Walter, Research and Development Analyst for Nielsen Bookscan, for his generosity in supplying these sales figures.

Kinsella), romance (Danielle Steel), women's fiction (Maeve Binchy, Jodi Picoult), or historical novels (Bernard Cornwell, Sebastian Faulks). Several issues arise from this initial approach: on the one hand, the sub-categories we mention here are not comprehensively discrete, that is, there are no fool-proof criteria to define these sub-genres as unequivocally distinctive from one another. For instance, the mystery and suspense, and thriller sub-groups, on the one hand, and the Chick Lit and romance, on the other, could be said to share some important narrative and structural characteristics and should be, perhaps, best viewed as part of a continuum rather than as totally independent forms. Another example is the category 'women's fiction' which is not easily pinned down, nor do all critics use it to refer to the same type of writing. Whereas some authors might employ it as a kind of umbrella term to encompass all forms of writing by women, others prefer to restrict the sense in which the label is employed to specifically describe literary fiction and to distinguish it categorically from other forms of popular writing such as Chick Lit or romances. Moreover, some of the works by the authors above could actually be rehoused into categories other than the ones we assign them to; in sum, this initial attempt at a definition of 'popular-ness' demonstrates that this is a complex task⁶.

If, as McCracken suggests above, popular forms are mainly characterised by their large readership, then all the authors above should be called popular writers. But McCracken also adds that it would be over-simplistic to think simply in terms of numbers, and that 'in the context of the late twentieth century, that definition needs some refinement' (1998: 1). McCracken further suggests that popular fiction needs to be understood as part of an 'entertainment industry' which includes 'popular narratives for film, radio, television and periodicals as well as in book form' (1998: 1). This highlights both the multitude of forms into which the notion of 'popular-ness' can materialise and the fact that popular forms are highly motivated by a desire to entertain. We come back below to this idea of how entertainment is associated with popular fiction and we also look into how the entertainment aspects and the lack of complexity which readers and literary critics appear to associate with genre fiction are two issues directly linked.

Gelder (2004), in turn, defines popular fiction as follows:

Popular fiction is best conceived as the opposite of Literature (to which I shall ascribe a capital L, distinguishing it from literature as a general field of writing). [...] By Literature, I mean the kind of writing [...] produced by, for example, Jane Austen, George Elliot, Henry James, James Joyce, William Faulkner [...] Saul Bellow, D.H. Lawrence, Flannery O'Connor, Vladimir Nabokov, Martin Amis [...] and so on. (Gelder, 2004: 11)

Gelder starts off by highlighting the (apparently) antithetical relationship between popular and serious forms in so much as the former is clearly what the latter is not. Thus, Literature, with capital 'L' (what we refer to as serious/literary fiction) is exemplified by writers who could be said to illustrate the canon of the 19th and 20th century literary landscape in the English language. Gelder, however, fails to provide any criteria according to which the above authors should be considered prototypically 'Literary'. He instead complements his appraisal by also referring to readership

⁶ A most notable example of an author who has bridged the popular vs. serious distinction is Charles Dickens.

numbers as an important factor, but he appears less convinced that this is crucial to distinguish literary from popular fiction:

Some of these writers may even have written what could be termed 'best-sellers', although this term is quantitatively open: a bestseller can mean sales of anything from around 20,000 copies to several million (after which, we might use the terms 'superseller' or topseller), and some works of Literature, whether it happens over an extended period of time or immediately after publication, can indeed do well in the marketplace. (Gelder, 2004: 11)

Gelder echoes the distinction drawn by the French sociologist Pierre Bourdieu (1996) in relation to literary and genre fiction, which he summarises as follows:

Bourdieu characterizes high or highbrow cultural production⁷ [...] as 'autonomous': indifferent to the buying and reading/viewing public [...], underwritten by a sense of 'creativity' and 'originality', and using the language or discourse of 'art'. [...] By contrast, a form of low cultural production such as soap opera is described by Bourdieu as 'heteronomous': open to mass audiences and necessarily caught up in the logic of the market place [...]. It usually doesn't draw on the language of the art to define itself; more commonly, it uses the language of industry and production instead. (Gelder, 2004: 13)

Our interest in investigating the linguistic characteristics of popular fiction in relation to serious fiction stems directly from appraisals such as Gelder's and Bourdieu's above which, though valid as a way to account for the socio-cultural differences between the two forms, fail to describe how these text varieties differ in linguistic terms. For instance, as McCracken also suggests above, one initial conclusion that can be drawn concerning the difference between popular and serious forms is the fact that the former pay heed to marketability concerns much more clearly than the latter. Bourdieu (1996) defines this particular attention to audience/readership demands in terms of the autonomy and heteronomy concerns which typify serious and popular fiction respectively. Gelder, then, suggests (alongside Bourdieu) that this autonomy is achieved by 'using the language or discourse of art' (2004: 13) whereas the heteronomy which prototypically characterises lowbrow fictional manifestations 'doesn't draw on the language of the art to define itself; more commonly, it uses the language of industry and production instead' (2004: 13). Gelder goes on to claim that by using the language of art, literary fiction has "intense formal artistry" and "tangled plots" which perfectly capture the "tangled lives" found in the ordinary outside world' (2004: 13) whereas popular fiction is better off dispensing with such a technique: 'for popular fiction, however, "tangled plots" would be a damning criticism, rather than something to be proud of – and as for "intense formal artistry", perhaps the less said the better' (Gelder 2004: 13). Gelder, though, does not elaborate on how to understand 'the language of art' or 'the language of industry and production', nor does he explain why employing tangled plots to capture the tangled lives of the outside world would be such an alien or damning characteristic for popular fiction. Gelder's evaluation, then, falls short of providing objective linguistic and stylistic parameters to distinguish these two text varieties.

⁷ Just as MacCracken (1998) does above, Bourdieu also sees fiction as a cultural artefact part of a much wider enterprise which he refers to as 'cultural production'.

Some other attempts emanating from the literary critical world have come slightly closer to using linguistic models or, at least, linguistic terminology, to achieve their aim, as illustrated here by Moretti:

Prose is not a gift; it's work: [...] hypotaxis is not only laborious—it requires foresight, memory, adequation of means to ends—but truly productive: the outcome is usually more than the sum of its parts because subordination establishes a hierarchy among clauses, meaning becomes articulated, aspects emerge that didn't exist before. That's how complexity comes into being. (Moretti, 2010: 1)

Moretti (2010), admittedly, is not concerned with accounting for the differences between high and low-brow fiction. His main focus is on the history and theory of the novel, but the fact that he bases his appraisal on linguistic factors (albeit in a rather rudimentary way) is relevant to our interests for even the literary critical establishment would appear to acknowledge that a linguistic/stylistic characterisation is of benefit to the endeavour that occupies us here. Furthermore, it is encouraging that voices emanating from the literary critical world appreciate the usefulness of incorporating digitalised data and corpus methodologies for the analysis of literature:

Looking at prose style from below with digital databases, this is now easy to imagine: a few years, and we'll be able to search just about all novels that have ever been published and look for patterns among billions of sentences. Personally, I am fascinated by this encounter of the formal and the quantitative (Moretti, 2010: 3).

Literary criticism has also focused on the characteristics of specific popular authors. For instance, Ramet (1999) analyses the work of the writer Ken Follett⁸, specifically *The Modigliani Scandal* (1976):

Obviously, much of this still depends on summary narration and on the compactness noticeable in journalism. Looking at the paragraphs more scrupulously, we can observe that this unadorned style is the result of two simple sentences of ten words and eight words which begin the passage, and then are followed by three sentences of sixteen words each. Those three subsequent sentences, however, each has the feel of much shorter sentences since all three remain compound sentences reliant on the coordinating conjunction 'and'. The slight variation in the last sentence is achieved by 'and' once in the compound predicate. (Ramet, 1999: 13)

At first glance, Ramet's analysis seems much closer to the kind of linguistic evaluation we pursue in our own research. On closer inspection, though, some of the remarks appear unfounded and not totally justified from a linguistic perspective. For instance, Ramet characterises some passages in Follett's work as journalistic because of the word-per-sentence ratio; somehow, he concludes that both the first two simple sentences containing ten and eight words respectively and the compound sentences that ensue (sixteen words each) illustrate the kind of register prototypically used in

⁸ Related to the marketability issue discussed previously, Follett's work also features in the Nielsen Bookscan top 100 list in 44th position.

journalism. It does not seem immediately obvious, though, how he ends up likening the word-per-sentence ratio to a particular register; nor is it obvious why the final three sentences might 'feel' shorter simply because they are compound sentences. The following is a further illustration of the kind of, otherwise sound, grammatical analysis produced by Ramet (1999) which, however, fails to yield fruitful results with regard to the effects created by the use of such devices:

Paper Money (1977), written the following year and about an ambitious young reporter for a London newspaper, is stylistically still reliant on simple sentences, resulting in what one reviewer termed 'a quick and entertaining read'. But these simple sentences are now balanced out by many more concrete nouns, complex sentences, and subordinate clauses. (Ramet, 1999: 13)

There is no evidence above as to why concrete nouns, complex sentences and subordinate clauses would 'balance out' simple sentences, nor why, in fact, simple sentences would need to be balanced by any other grammatical unit. If, as one reviewer suggests, simple sentences create a 'quick and entertaining read', might we conclude that literary texts are all 'slow and unentertaining' on account of their apparent complexity? Furthermore, in the above quotation Ramet analyses *Paper Money* (1977), a novel Follett published just one year after *The Modigliani Scandal* (1976) described in the previous quotation. In the space of one year, Follett seems to have mastered a certain stylistic maturity by veering away from the journalistic style created through a series of simple sentences towards a more complex style projected by a combination of simple sentences balanced out by concrete nouns and certain other linguistic devices.

It is, therefore, this lax methodological approach that we want to address. Our initial tenets are that the claims by literary critics tend to be rather too imprecise to be testable. Often, their claims (though not necessarily wrong) are simply made on the basis of very little evidence and this lack of evidence reduces their usefulness as hypotheses. Two specific claims can be extrapolated from the general views proposed so far:

- Claim 1: Literature is different from popular fiction; Literature, to quote Gelder (2004: 13), 'draws on the language of the art world'⁹. This fact distinguishes Literature from popular writing which is, by implication, not reliant on or directly informed by the discourse of the art world. Literature, therefore, is seemingly considered intrinsically 'artistic', whereas popular forms are not.
- Claim 2: Syntactic complexity is seen as a stylistic development; that is, the more formally complex a text is, the closer to Literature and, consequently, also to that art world mentioned above.

Our aim is to address these claims from a corpus stylistic perspective to ascertain whether they can indeed be validated. Additionally, our analysis allows us to reflect upon the notion of syntactic complexity being seen as a hallmark of good writing not just by literary critics but also by non-academic readers. As Mahlberg (2013: 2) states: 'A corpus stylistic approach assumes that the linguistic analysis of a literary text provides useful insights complementing the literary appreciation of the work. [...] Corpus stylistics [...] draws on corpus methodology but at the same time, it emphasises

⁹ Note that this too is a rather vague claim that could be tested through a lexical and syntactic investigation of a corpus of texts drawn from the art world (exhibition catalogues, for example).

the link that literary stylistics provides to literary criticism'. Thus, just as Mahlberg contends here, we take on board literary critical concerns with regard to the difference between serious and popular fiction but argue, unlike most literary critics, that our understanding of the serious vs. popular dichotomy can be augmented by adopting the methodical and rigorous approaches on which corpus linguistics relies. In sum, we argue that corpus stylistics, as a discipline, is better suited to account for which testable and replicable criteria distinguish one text type from the other.

3. Data and method

As described in the introduction, we analyse two target corpora, the Lancaster Speech, Writing and Thought presentation corpus (SW&TP) (Semino and Short, 2004) and what we have called *The Chick Lit Corpus (CLC)* (Montoro, 2012). The former contains 80,000 words roughly divided into two sub-corpora of 40,000 words each, representing serious and popular fiction respectively. The corpus was created at Lancaster University and is described by Semino and Short (2004) as follows:

Our corpus contains 120 text samples of approximately 2,000 words each, amounting to a total of 258,348 words of (late) twentieth-century written British English. It is divided into three sections, which comprise 40 text samples each and represent three main genres:

- Prose fiction (87,709 words)
- Newspaper news reports (83,603 words)
- Biography and autobiography (87,036 words)

(Semino and Short, 2004: 19)

Each of the three sections of the Lancaster corpus was subsequently subdivided into 'serious' and 'popular' subsections according to a series of parameters although the authors admit that this division was not free from controversy. The criteria used for the distinction, especially as far as the fiction section of the corpus is concerned, are, for the most part, not linguistic, but 'the cultural distinction between popular fiction on the one hand and highbrow fiction on the other. Apart from anything else, this distinction has concrete manifestations, such as the design of front covers and the physical positioning of books in bookshops and libraries' (Semino and Short, 2004: 22). Half the samples they used were selected from the *Oxford Text Archive (OTA)* and the *British National Corpus (BNC)*; the second half were texts published contemporaneously to the time when the authors were creating the corpus (late 20th century). In the serious sub-set of samples we find authors such as Virginia Woolf, D. H. Lawrence and Salman Rushdie; in the popular sub-section are Catherine Cookson, Rupert Thomson and Wilbur Smith, among others.

Due to the lack of clear-cut and agreed upon criteria to separate these two text types, we were particularly concerned with the circularity problem in our own research; that is, we were aware that the validity of the results obtained would be weakened, if not totally invalidated, if the classification of data as either serious or popular was decided only by us. The fact that the SW&TP was created by a separate team of analysts provided the impartiality we were after, irrespective of whether their criteria are correct or fully agreed on by all scholars. The choice of our second corpus, the *CLC*, also responded to an interest to maintain an unbiased approach to the material analysed. Thus, we considered a relatively recent example of popular genre, Chick Lit, as already defined by scholars such as Ferriss and Young (2006), Gormley (2009) and Whelehan

(2002). Based on their set of criteria for the identification of Chick Lit works, inclusive of particular book covers, female authors and female readers, we used the *CLC* (Montoro, 2012), which is made up of six novels amounting to a total of about 500,000 words¹⁰. With this second corpus, we aimed to expand the analysis implemented by looking at the SW&TP to further test our initial hypothesis.

After the selection of corpora, the next stage was to focus on the one variable we were concerned about, namely subordinating conjunctions. As explained above, we based our decision on the repeated claim that serious fiction is structurally complex, an assessment which often gets equated with elaboration projected mainly by subordination. While subordination can be marked in other ways than through the use of subordinating conjunctions (see note 2), we chose to use these as an indicator of subordination simply because this form-based approach lends itself to automatic corpus linguistic analysis. To locate the subordinating conjunctions, we tagged both the SW&TP and the *CLC* for POS information using the CLAWS tagger in Wmatrix (Rayson, 2009). The third stage in our investigation was to decide on the most appropriate reference corpora. We were motivated primarily by thoroughness and comparability¹¹ and the software Wmatrix (Rayson, 2009) enabled us to satisfy both criteria. In order to meet the first criterion, we compared the SW&TP and the *CLC* to the circa 1,000,000-word sample of the BNC. For the second criterion, we viewed our corpora in relation to the circa 225,000-word BNC Written Imaginative sub-set which contains only fiction. Based on Mandala's (2010) and Stockwell's (2000) conclusions with regard to the complexity of certain popular writing varieties when compared to spoken language, we finally compared the speech presentation sub-sets of the SW&TP to the BNC Spoken. The various sets of comparisons are summarised in Table 1:

Target Corpora	Reference corpora		
	BNC Written Sampler (c. 1,000,000 words)	BNC Written Imaginative (c. 225,000 words)	BNC Spoken (c. 1,000,000 words)
SW&TP – Serious (c. 40,000 words)	√	√	
SW&TP – Popular (c. 40,000 words)	√	√	
SW&TP – Serious Narration only (c. 20,000 words)	√	√	
SW&TP – Popular Narration only (c. 19,000 words)	√	√	
SW&TP – Serious Speech presentation only (c. 23,000 words)			√

¹⁰ The six novels are *Bridget Jones's Diary* (Fielding, 1996), *Getting Rid of Matthew* (Fallon, 2007), *Pants on Fire* (Alderson, 2000), *The Last Year of Being Single* (Tucker, 2003), *The Other Woman* (Green, 2004), and *The Secret Dreamworld of a Shopaholic* (Kinsella, 2000).

¹¹ The issue of corpora comparability is one of the most discussed and sometimes contentious aspects in corpus linguistics. See McEnery and Hardie (2012) for more on this.

SW&TP – Popular Speech presentation only (c. 25,000 words)			√
Chick Lit Corpus (CLC) (c. 500,000 words)	√	√	

Table 1 Target and reference corpora: Implemented comparisons

To measure the statistical significance of the outcome of such comparisons, Wmatrix uses the log-likelihood test (Dunning, 1993). We set an alpha level of 0.05 (LL critical value = 3.84). This decision was based on 0.05 being a generally accepted significance level within the social sciences. To reduce the possibility of a type I error, we could have lowered the significance level further, though we decided not to do this so as not to exclude results which may have fallen just below the threshold for significance. Ultimately, the keyness analysis enabled us to check whether there is indeed more syntactic complexity in serious fiction and whether this can be said to be realised by subordinating conjunctions.

A second advantage of using the SW&TP corpus is the fact that this corpus has been previously annotated for Speech, Writing and Thought presentation categories. Indeed, the motivation behind the construction of the SW&TP corpus was to test empirically the Speech and Thought presentation model proposed by Leech and Short ([1981] 2007). Semino and Short (2004) (together with a team of analysts) annotated the SW&TP manually to incorporate phenomena that works at discoursal level and which were not easily identifiable by the existing software. The text mark-up was done by using a set of SGML-conformant conventions inclusive of text division, text boundaries, page breaks and the corresponding 'sptag' (that is, speech, writing and thought presentation categories)¹². The latter tags allowed us to isolate and extract stretches of text inclusive of speech presentation and to separate them from narration. This separation enabled us to exploit this corpus in two different ways. On the one hand, we looked at the corpus as a target corpus in its own right since it contains examples of 20th century serious and popular fiction; on the other, having isolated instances of speech from stretches of narration we were able to consider the way syntactic complexity is employed in each prose variant respectively. In what follows we describe in detail the results of the set of comparisons we implemented and, more importantly, discuss the stylistic effects of the over- or under-use of subordinating conjunctions, first, followed by a series of additional POS tags which our corpus-based approach brought to the fore and which literary critical assessments had, however, failed to address.

¹² The annotation of the corpus includes information on the SW&TP category of the particular stretch analysed, the SW&TP category of the immediately following text ('next'), etc. (Semino and Short, 2004: 29). The resulting tagged corpus looks as follows:

`<sptag cat=NRS>The Political Officer at Vladimir said</sptag><sptag cat=IS>that it would not be too long, </sptag><sptag cat=NRSA>and the interrogations had been courteous`

The sptag cat=NRS represents the Narrator's Report of Speech, the sptag cat=IS stands for Indirect Speech and sptag cat=NRSA refers to the Narrator's Representation of Speech Acts; for the full set of categories see Semino and Short (2004: 30-31).

4. Results and discussion

The following tables contain the results of tagging our target corpora for POS information and comparing them with the various reference corpora (summarised in Table 1). Tables 2-8 include information exclusively on the variable which is the main (though not the only) concern of this paper, that is, subordinating conjunctions (indicated by the POS tag 'CS'). We start by comparing the serious and popular fiction sub-sets of the SW&TP corpus with the BNC written sampler (Tables 2 and 3):

	Serious fiction	Observed frequency %	BNC written sampler	Observed frequency %	Log-likelihood score
CS21 ¹³	61	0.15	659	0.07	+26.14
CS	393	0.94	7380	0.76	+14.76
CSA (<i>as</i> (as conjunction))	153	0.36	2707	0.28	+9.38
CS31	10	0.02	100	0.01	+5.08

Table 2 Serious fiction compared with BNC written sampler ($p < 0.05$; LL critical value = 3.84)

	Popular fiction	Observed frequency %	BNC written sampler	Observed frequency %	Log-likelihood score
CS	355	0.94	7380	0.76	+13.86
CS32	1	0.00	0	0.00	+6.56
CS31	10	0.03	100	0.01	+6.27

Table 3 Popular fiction compared with BNC written sampler ($p < 0.05$; LL critical value = 3.84)

As the results in Tables 2 and 3 demonstrate, the subordinating conjunctions are statistically over-used in both popular and serious fiction. On this basis alone, there is no evidence to support the popularly held notion that popular fiction is less complex syntactically than serious fiction.¹⁴ In order to test whether the over-use of subordinating conjunctions was influenced by the reference corpus, we compared both the serious and popular sub-corpora with the BNC Written Imaginative and obtained similar results. In the former, the CS21 tag was also over-used (LL = + 11.62) as well as the CST tag (*that* (as conjunction); LL = + 9.44); in the latter, besides the three-part subordinator CS31 (see below) also over-represented (LL = +3.86), we encountered the CST tag (*that* (as conjunction); LL = + 9.43) and the CSW tag (*whether* (as conjunction); LL = + 5.48). What is interesting about these subordinating conjunctions, however, is their function and the meaning they help project in the two text varieties they are found in, particularly as far as the two (CS21) and three-part subordinators (CS31) are concerned. In the SW&TP serious fiction sub-corpus, there are 61 instances

¹³ The first of the two digits immediately following the grammatical tag 'CS' (as well as any other grammatical tag, as will be seen below) indicate the number of words/tags in the sequence; the second digit marks the position within that sequence. For instance, *as if, so that, now that, as though*, discussed below.

¹⁴ N.B. The zero score in Table 3 for the observed frequency of CS32 in the target and reference corpora is an indication of the very small percentage of the corpora that is composed of this particular POS.

of the two-part subordinator CS21, realised by the following conjunctions (also known as complex subordinators (Biber et al., 1999: 85)) and with the following frequencies:

AS IF (30)
 SO THAT (10)
 NOW THAT (7)
 AS THOUGH (6)
 IN CASE (2)
 EVEN IF (2)
 EVEN THOUGH (2)
 SAVE THAT (1)
 RATHER THAN (1)

Consider the following concordance lines for the top-scorer within the CS21 category, namely *as if*:

d not wish to leave the quarterdeck	as if	retreating from it . "Cumbershum
hy, even they stood and watched	as if	someone had said hist, while the
ONE RANG AGAIN . JIM TATE LOOKED	as if	IT MIGHT STING HIM . "YOU'RE BU
ONLY WOULD . IF YOU 'D SORT OF LOOK	as if	YOU WAS ORDERING THINGS . THERE 'S N
WALKED WITHOUT A STICK. " LOOKS	as if	HE WANTED TO TALK TO US , " SAID
LE PERSON SHE DETECTED A FLAME ;	as if	A SPARK IN THE BRAIN IGNITED SPON
that he devoured . now and then ,	as if	recapturing his lost childhood , jim
r gave a diagonal nod of the head ,	as if	it were slightly impertinent to know

Figure 1 Concordance for POS tag CS21 (*as if*) - Serious Fiction

One preliminary conclusion that could be drawn from these examples is the fact that the complex sentences in which the *as if* clauses appear seem to be providing, for the most part, simile comparisons (e.g. 'as if a spark in the brain') or appear to be used to reformulate (by comparing) a previous proposition (e.g. '... gave a diagonal nod of the head, as if it were slightly impertinent to ...')¹⁵. It is often the case that the *as if* clause is an instance of internal narration (NI¹⁶ in SW&TP terms), which supports the idea that literary fiction tends to be more reflective and contemplative than popular fiction. By contrast, there are no instances of the CS21 category (hence, no *as if* constructions) in popular fiction. In popular fiction, besides the CS category, we also find the CS31 tag, a three-part subordinator realised by the conjunctions *as soon as*, *as far as*, and *as long as*. Interestingly, this ties in with the notion that popular fiction is more concerned with action than with reflection, as prototypically found in thrillers, crime fiction or science fiction. The following examples illustrate the CS31 tag in popular fiction:

it was only a coincidence but ,	as soon as	Jed pronounced the name of his em
ring her holiday in France . But	as soon as	it was repeated , its individuality
very firm too . You are innocent	as far as	Her Majesty 's Government is conce
our folly , we started pushing it	as far as	it would go . This was made perfec
en said doubtfully: " Well ,	as long as	he doesn't alter course . "
's a great source , is Dennis ,	as long as	the people I 'm after have no con

Figure 2 Concordance for POS tag CS31 (*as soon as*, *as far as*, *as long as*) - Popular Fiction

¹⁵ On the role and function of *as if* clauses, see also Mahlberg (2013).

¹⁶ Internal narration refers to 'those cases where the narrator reports a character's cognitive and emotional experiences without presenting any specific thoughts' (Semino and Short, 2004: 46)

The introspection and comparative structures characterising some of the concordance lines in Figure 1 are not observable in the examples above, where, for instance, the subordinator *as soon as* (the most recurrent realisation of CS31) presents, instead, one action in relation to another. Due to the lack of evidence so far, mainly related to the relatively small size of the SW&TP corpora, we refrain from making over-ambitious claims with regard to the generalisability of our conclusions here. However, these results seem to suggest that it is not the lack of syntactic complexity generated by the under-use of subordinating conjunctions but the functions these conjunctions have that help identify each text variety as distinct from each other. That is, it cannot be said that popular fiction lacks complexity because subordinating conjunctions are the dis-preferred choice; popular writing simply appears to use those conjunctions that facilitate the expression of action avoiding the reflective nature of prototypical serious fiction. We elaborate on this notion further when we discuss the *CLC* below.

The next stage in our analysis was to consider the speech and narration sections of the SW&TP. Our second aim was to ascertain whether the syntactic complexity we investigate could be explained not by the dichotomy serious vs. popular fiction but by stylistic techniques such as the presentation of the narrator’s voice vs. characters’ speech. The way the SW&TP was constructed and manually annotated made it relatively straightforward to isolate and separate speech from narrative sections (by means of the ‘sptags’ described above). So, we approach this next premise by first comparing each sub-set of the SW&TP with the BNC Written sampler and then by viewing them with regard to the BNC Written Imaginative.

	Serious fiction (narration only)	Observed frequency %	BNC written	Observed frequency %	Log-likelihood score
CS21	33	0.18	659	0.07	+22.21
CSA (<i>as</i> (as conjunction))	80	0.43	2707	0.28	+13.00
CS	167	0.90	7380	0.76	+4.49

Table 4 Serious Fiction (narration only) compared with BNC Written sampler ($p < 0.05$; LL critical value = 3.84)

As Table 4 illustrates, subordinating conjunctions are also over-represented in the narration only sections of the SW&TP’s serious sub-set. Moreover, we find that the same CS21 tag (*as if*, for instance) is also key. The results in Table 5 below confirm that subordinating conjunctions, including the three-part subordinator CS31 (*as soon as*, *as far as*, *as long as*), are over-represented in the narration only sub-set of the popular fiction corpus.

	Popular fiction (narration only)	Observed frequency %	BNC written	Observed frequency %	Log-likelihood score
CS31	6	0.04	100	0.01	+6.42
CSA (<i>as</i> (as conjunction))	62	0.38	2707	0.28	+5.16

Table 5 Popular Fiction (narration only) compared to BNC Written sampler ($p < 0.05$; LL critical value = 3.84)

The statistical significance of the results in Tables 4 and 5 is further supported by the results of comparing the narration only sections of both corpora with the BNC Written Imaginative. In the narrative section of the SW&TP’s serious sub-set the CS21 (LL = +12.70) and the CSA (LL = +7.08) tags are over-represented; in the narration only section of the SW&TP’s popular sub-set the CSW31 tag is over-used (LL = +5.47). Based on these results, subordinating conjunctions are key (that is, over-represented) in both serious and popular fiction, which would indicate that the narrative voice in popular writing is as syntactically complex (that is, realised by subordinating conjunctions) as it is in literary fiction.

Mandala (2010) focuses on the linguistic characteristics of two prototypically popular narrative genres: science fiction and fantasy. Based on Taylor (1990) and Stockwell (2000), she puts together a list of what she terms ‘pulp features’:

Intrusive and over-long passages of exposition; story titles that resemble newspaper headlines with their use of short and simple noun phrases (e.g. ‘Big Game’); needless repetition of adjectives and adverbs that serve as little more than fillers; a dependence on third person omniscient narration; dialogue that is too complex syntactically to convincingly represent talk; and a tendency to use unnatural synonyms for the speech reporting verb ‘said’, and to modify those synonyms with equally unnatural adverbs (e.g. ‘No!’, he blustered roarily). (Mandala, 2010: 16-17)

Although we cannot discuss this list in detail, we are particularly interested in Mandala’s characterisation of pulp dialogue as ‘too complex syntactically to convincingly represent talk’ (Mandala, 2010: 17). This assessment introduces a new element not considered by some of the literary critical evaluations we discuss in section 2, whereby the syntactic complexity of popular dialogue is measured against the features of spoken language¹⁷. Consequently, we next tested the speech presentation only sub-sets of the SW&TP against the BNC spoken corpus (in Tables 6 and 7). Our aim was to consider whether the dichotomy serious vs. popular might be determined by the alignment and/or distinction of either text type with spoken forms:

	Serious fiction (speech presentation only)	Observed frequency %	BNC spoken	Observed frequency %	Log-likelihood score
CSA (<i>as</i> (as conjunction))	69	0.30	1335	0.14	+33.68
CS21	29	0.13	555	0.06	+14.50
CSN (<i>than</i> (as conjunction))	29	0.13	653	0.07	+9.77
CST (<i>that</i> (as conjunction))	192	0.84	6644	0.68	+8.66

¹⁷ On subordination and spoken language see Barth (2000), Couper-Kuhlen and Thompson (2000) and Ford (1993).

Table 6 Serious Fiction (speech presentation only) compared with BNC Spoken sampler ($p < 0.05$; LL critical value = 3.84)

	Popular fiction (speech presentation only)	Observed frequency %	BNC spoken	Observed frequency %	Log-likelihood score
CSA (<i>as</i> (as conjunction))	73	0.28	1335	0.14	+28.51
CST (<i>that</i> (as conjunction))	225	0.85	6644	0.68	+10.89
CSN (<i>than</i> (as conjunction))	27	0.10	653	0.07	+4.24

Table 7 Popular Fiction (speech presentation only) compared with BNC Spoken sampler ($p < 0.05$; LL critical value = 3.84)

As both tables make clear, subordinating conjunctions are over-used in the speech presentation sub-sets of the serious and popular sub-corpora when compared with spoken language. These results suggest that fictional dialogue, irrespective of its high or lowbrow status is more syntactically complex (in the sense defined in this paper) than spoken language. Biber et al. (1999) address the frequency and distribution of subordinating conjunctions in conversation and fiction:

It is somewhat surprising that subordinators are relatively rare in all registers, but somewhat more common in conversation and fiction than in academic prose and news reportage. (Biber et al., 1999: 93)

If the above is the case, two, albeit tentative, conclusions can be drawn from the assessments proposed by Mandala (2010) and Biber et al. (1999). Firstly, following on from the latter's (1999) results, it could be argued that, not only are subordinators more prevalent in conversation and fiction, but also, that the latter uses some of these subordinating conjunctions more frequently, which makes these two registers distinct from each other with regard to this particular aspect. Secondly, although Mandala does not explicitly discuss syntactic complexity realised by subordinating conjunctions, our results seem germane to qualify her claim by specifying with respect to what POS the syntactic complexity she underscores in popular fiction appears to work, that is, via the over-use of subordinating conjunctions. Our results, therefore, seem useful to fine-tune previous assessments and to suggest ways forward to investigate this issue further as we now do by turning our attention to one specific sub-genre of popular writing, Chick Lit.

Chick Lit has received some attention from literary scholars (Ferriss and Young, 2006; Whelehan, 2002) who prototypically align it with popular fiction. Montoro (2012) examines a corpus she constructed containing the unprompted online opinions on the genre published in specifically devoted blogs and book-club forums where readers explain exactly what they consider to be the real appeal of the genre:

Karen (27th March 2008)

I have to admit to being a chick-lit fan and I'm proud of it. I know they're **not the most intellectual books** on the market but I like them because **they're easy to read** and more often than not are fluffy and light-hearted

Purple Princess (17th August 2009)

Have finished *Twenties Girl* and I thoroughly enjoyed it! The characters grew on me as I got further through the book and although it is a bit farfetched I still really like **her style of writing and it was just so easy to read.**

Inver (29th October 2009)

Recently lent this one to a fellow book-crossing friend and she said....

Quite **complex themes dealt with in a way that**, once you had worked out who was who, **made it very readable.** In fact, once I got into the book, I found it hard to put down.

Chick Lit readers characterise this genre as 'easy to read' and, importantly, 'easy-ness' represents precisely one of the main appeals of these works which are felt to be entertaining precisely because they do not pose strenuous linguistic challenges. Thus, the 'easier' the book, the more appealing it becomes. Linguistic simplicity, therefore, would appear to be as much an expectation held by the readership as a defining characteristic of the genre often picked up by critics. In order to develop our initial analysis, we next investigate whether the linguistic simplicity of Chick Lit expected by readers can be statistically supported by the under-representation of subordinating conjunctions. As implemented above, we compare our target corpus, this time the *CLC*, with the BNC Written Sampler and the BNC Written Imaginative sub-corpora. Briefly, our results corroborate that subordinating conjunctions are statistically over-represented, so this particular manifestation of low-brow literature conforms to the trends ascertained in our analysis above:

<i>CLC</i>	BNC Written LL value	BNC Written Imaginative LL value
CS	+268.37	+37.56
CS21	+159.23	+35.92
CSW (<i>whether</i> (as conjunction))	+57.54	+21.00
CST (<i>that</i> (as conjunction))	No instances	+17.41
CS31	+43.38	+6.65
CSA (<i>as</i> (as conjunction))	+28.89	No instances
CS22	+4.03	No instances

Table 8 Subordinating conjunctions in *CLC* ($p < 0.05$; LL critical value = 3.84)

Table 8 summarises the number of CS tags over-represented in the *CLC* when compared to the BNC Written and the BNC Written Imaginative. From these results, it seems that syntactic simplicity is not a feature of Chick Lit. Additionally, it is interesting to see the kind of CS tags used in the *CLC* as both CS21 (*as if, as though, even though*) and CS31

(*as soon as, as long as, as far as*) are statistically and interpretatively salient. Unlike the sample texts contained in the popular sub-section of the SW&TP, Chick Lit novels seem capable of displaying both the reflection associated with literary forms and the action linked with popular fiction. These results suggest that popular writing is far more stylistically intricate and shows more linguistic versatility than initially considered and also points out that the narrative and stylistic variety within the general umbrella term of popular fiction is a further factor to consider. Defining popular writing as a less sophisticated, less developed manifestation of fiction, therefore, is a concept that corpus stylistics seems capable of challenging and claiming as erroneous.

What the results in Table 8 did indicate, though, was that a reformulation and expansion of our initial hypothesis was necessary and that it could, in fact, yield fruitful results. As quoted above, Biber et al. (1999: 93) account for the frequency and distribution of subordinating conjunctions with regard to register and conclude that 'subordinators are relatively rare in all registers, but somewhat more common in conversation and fiction'. Additionally, they argue that in certain registers, complexity is realised at phrasal rather than clausal level:

This distribution reflects the fact that complexity in expository writing resides at the phrase level rather than at the level of clause combinations (thus the high frequency of prepositions; [...]). In contrast, conversation and fiction have a high frequency of verbs, hence also of clauses and clause combinations. As differences are less marked with coordinators and subordinators than with the function words that operate specifically at the phrase level, it seems justified to conclude that register differences are more connected with the build-up of phrases than with the connection of clauses. (Biber et al., 1999: 93)

According to Biber et al. (1993) fiction, by virtue of its higher frequency of verbs, is more likely to display complexity at the clause and clause combination level. Register differences, though, seem to be explained by the 'build-up of phrases' so that the complexity of expository writing relies on its intricate accumulations at phrasal level. They conclude, therefore, that register distinction is highly reliant on the elaboration (or lack of it) at phrasal level. This prompted us to assess whether linguistic simplicity working at phrasal level was indeed a feature of popular writing. Thus we extended our analysis to consider not simply linguistic complexity realised by the over-use of certain forms, but to investigate whether, in fact, a correlate 'lack' of complexity was projected by the under-use of other POS. Literary critics have neglected to pay attention to this different but complementary set of key results which, crucially, appear to account more accurately for the prototypical linguistic make-up of popular fiction. In Table 9, we focus on three under-used POS in the *CLC* (compared with the *BNC Written* and the *BNC Written Imaginative*; not all the under-used POS are included nor discussed): prepositions, in particular, *of*; adjectives, and the *wh-* determiner, *whose*:

<i>CLC</i>	BNC Written LL value	BNC Written Imaginative LL value
IO (<i>of</i> , as preposition)	-5109.16	-264.68
JJ (general adjective)	-2142.91	n/a
II (general preposition)	-1770.60	-47.65

JJR (general comparative adjective (e.g. older, better, stronger))	-162.66	-6.57
II31	-97.12	n/a
II33	-77.47	n/a
DDQGE (wh-determiner, genitive [whose])	-67.55	-12.04
JJ22	-18.10	-28.80
II41	-15.96	n/a
JJ21	-10.15	-14.10
II22	n/a	-10.09
JJT (general superlative adjective (e.g. oldest, best, strongest))	-7.56	n/a

Table 9 Under-used POS in *CLC* compared with BNC Written and BNC Written Imaginative ($p < 0.05$; LL critical value = 3.84)

The IO tag (*of*, as preposition) is in the top 10 most significantly under-represented categories in the *CLC* (conversely, the IO tag is over-used (LL = +103.25) in the serious sub-corpus of the SW&TP). On closer inspection, concordances show that the IO tags are mainly (though not exclusively) used to introduce post-modifying prepositional phrases in the NP¹⁸. For instance, consider the following concordance lines:

ity and sense of self as woman	of	substance , complete without bo
a week. Reduce circumference	of	thighs by 3 inches (e. 1 1/2
encial. Save up money in form	of	savings. Poss start pension al
ssical music . Give proportion	of	earnings to charity. Be kinder

Figure 3 Concordance for IO (*of* as preposition)

This under-use indicates that a lack of nominal post-modification could explain the stylistic simplicity prototypically seen as characteristic of popular fiction, at least with regard to Chick Lit novels. Elsewhere in the corpus, under-represented prepositions occur in prepositional phrases which function, among other things, as adverbials, as illustrated in Figure 4:

¹⁸ Montoro (2018 fc) expands on these issues by looking at an additional corpus (a c. 1,000,000-word corpus of modern vampire literature). Her research first corroborates that CS tags are also prevalent in her extended corpus of popular writing. Next, her results highlight the under-representation of the IO tag functioning as a nominal post-modifier, which suggests that phrasal rather than clausal simplicity does indeed characterise popular fiction. Together with the under-use of nouns also attested in her corpora, she concludes that popular writing is at odds with the 'densification' trend (Leech et al., 2009: 249) associated with 20th century English which, as a result, endows popular forms with the kind of colloquial feel often likened to 'easy-ness' and 'under-sophistication' by readers and critics.

hors to put impressively	on	shelves; exotic underwear
iend . Behave sluttishly	around	the house, but instead
circumference of thighs	by	3 inches (i. e. 1 1/2 inc
straight away when wake up	in	mornings . Go to gym three

Figure 4 Concordance for II (general preposition)

Thus, adverbials formally realised as prepositional phrases also seem to be dis-preferred in the *CLC* which would strengthen the idea of phrasal rather than clausal simplicity.

A further category which requires attention is that of adjectives (JJ tag) especially as our findings appear to directly contradict some of Mandala's (2010) 'pulp features' quoted above: 'a needless repetition of adjectives [...] that serve as little more than fillers'. Contrary to Mandala's observations, in the *CLC*, novels dispense with adjectives in a statistically significant way which is particularly telling as this sub-genre has been treated dismissively by many scholars for its apparent mindless and inane treatment of women's issues due to the constant description of clothes and shoes (see Ezard (2001) and Lessing (2001)). If there are indeed discussions of clothes and shoes, these do not seem to be accompanied by an over-use of adjectival phrases. Consequently, the under-use of a further POS bolsters the idea that phrasal simplicity characterises more accurately the linguistic features of popular writing. Additionally, the JJ tag, that is general adjective, is over- (LL = + 187.87) not underused in the serious fiction sub-corpus of the SW&TP which serves as further evidence of where the distinction serious vs. popular actually lies.

Finally, the category 'DDQGE wh-determiner, genitive (*whose*)¹⁹' is also under-represented:

lastic surgery was done on women	whose	husbands had run off with a
the brightest children are the ones	whose	parents talk to them a lot, even
that stuff in Ally Smith . The girl	whose	boyfriend was paying. Surely she
overblown name for a company	whose	clients were exclusively British

Figure 5 Concordance for DDQGE (wh-determiner, genitive [*whose*])

This under-occurrence underscores the fact that phrasal, rather than clausal, simplicity prototypically identifies Chick Lit novels because *whose* generally functions as post-modifier in noun phrases. All these instances of under-representation seem to confirm Biber et al.'s (1999) findings concerning the crucial role of phrasal elaboration to justify claims of syntactic complexity or simplicity. In the *CLC*, the latter emerges as a real characteristic especially when we consider the cumulative effect of the various under-representations.

4. Conclusion

Our aim in this paper has been to test whether one of the claims made about the language of popular fiction in relation to Literature (with a capital L) is actually tenable when viewed from an empiricallinguistic perspective. Following on from literary critical assessments of literary and genre fiction, we tested one specific aspect which is recurrently mentioned as prototypically distinguishing the former from the latter: syntactic complexity. The notion of complexity is multifarious and multifunctional in

¹⁹ This category is also over-used in the SW&TP serious sub-corpus: LL = +9.27.

the sense of including syntactic, stylistic, narrative and cultural components, so is hardly reducible to just one syntactic component. In order to generate a testable and falsifiable hypothesis, however, we focused particularly on subordinating conjunctions as a marker of syntactic subordination. Despite the claims in the literary critical literature for a major difference in syntactic complexity between literary and popular fiction, our results show that the assumed greater syntactic complexity of literary fiction is ill-founded and simply not substantiated. We have been able to show that syntactic complexity as marked by subordinating conjunctions is also prevalent in popular fiction. Moreover, as a consequence of our corpus stylistic analysis, we have been able to refine the literary critical claims we tested. For instance, we have shown that the perceived simplicity of one particular variety of popular fiction, Chick-lit, cannot be a consequence of the underuse of subordinating conjunctions; rather, if simplicity is perceived, this is likely to be a result of the underuse of particular POS at a phrasal rather than clausal level. This article, then, contributes not just to studies of so-called literary language but also to those of register differences. Ultimately, what we have demonstrated is the capacity of corpus stylistics to falsify literary critical claims and, by so doing, to contribute to literary critical understanding of the source of readers' reactions to fiction.

References

- Alderson, M. 2000. *Pants on Fire*. London: Penguin.
- Barth, D. 2000. "'that's true, although not really, but still": Expressing concession in spoken English' in E. Couper-Kuhlen and B. Kortmann (eds) *Cause, Condition, Concession Contrast*, pp. 411-437. Berlin and New York: Mouton de Gruyter.
- Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finnegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow, Essex: Pearson Education.
- Bourdieu, P. 1996. *The Rules of Art: Genesis and Structure of the Literary Field* (trans. by Susan Emanuel). Cambridge: Polity Press.
- British National Corpus (BNC), available at: <https://corpus.byu.edu/bnc/> [last accessed 2nd March 2018]
- Burke, M. 2014. 'Rhetoric and poetics: The classical heritage to stylistics' in M. Burke (ed.) *The Routledge Handbook of Stylistics*, pp. 11-30. Abingdon and New York: Routledge.
- Couper-Kuhlen, E. and SA. Thompson. 2000. 'Concessive patterns in conversation' in E. Couper-Kuhlen and B. Kortmann (eds) *Cause, Condition, Concession Contrast*, pp. 381-410. Berlin and New York: Mouton de Gruyter.
- De Bourcier, S. 2017. "'They all sound like David Foster Wallace": Syntax and narrative in *Infinite Jest*, *Brief Interviews with Hideous Men*, *Oblivion* and *The Pale King*', *Orbit: A Journal of American Literature* 5 (1), pp. 1-30.
- Dunning, T. 1993. 'Accurate methods for the statistics of surprise and coincidence', *Computational Linguistics* 19 (1), pp 61-74.
- Ezard, J. 2001. 'Bainbridge tilts at 'chick lit' cult', *The Guardian*, 24th August, available at: <http://www.guardian.co.uk/uk/2001/aug/24/books.generalfiction/> [last accessed 11th April 2018]
- Fallon, J. 2007. *Getting Rid of Matthew*. London: Penguin.
- Ferriss, S. and M. Young (eds) 2006. *Chick Lit. The New Woman's Fiction*. New York and London: Routledge.
- Fielding, H. 1996. *Bridget Jones's Diary*. London: Picador.

- Ford, C. E. 1993. *Grammar in Interaction: Adverbial Clauses in American English Conversation*. Cambridge: Cambridge University Press.
- Gelder, K. 2004. *Popular Fiction: The Logics and Practices of a Literary Field*. London: Routledge.
- Green, J. 2004. *The Other Woman*. London: Penguin.
- Gormley, S. 2009. 'Introduction', Working Papers on the Web, 13, available at: <http://extra.shu.ac.uk/wpw/chicklit/gormley.html> [last accessed 9th March 2018]
- Kinsella, S. 2000. *The Secret Dreamworld of a Shopaholic*. London: Black Swan.
- Lancaster Speech, Writing & Thought Presentation Corpus (SW&TP), available at: <http://ota.ox.ac.uk/> [last accessed 10th April 2018]
- Leech, G. and M. Short. [1981] 2007. *Style in Fiction. A Linguistic Introduction to English Fictional Prose*. 2nd edn. Harlow: Pearson Longman.
- Leech, G., Hundt, M., Mair, Ch., and Smith, N. 2009. *Change in Contemporary English. A Grammatical Study*. Cambridge: Cambridge University Press.
- Lessing, D. 2001. 'Bainbridge denounces chick-lit as 'froth'', *The Guardian Online*, available at: <http://www.guardian.co.uk/books/2001/aug/23/bookerprize2001.bookerprize?NTCMP=SRCH> [last accessed 10th April 2018]
- Mahlberg, M. 2013. *Corpus Stylistics and Dickens's Fiction*. New York and London: Routledge.
- Mandala, S. 2010. *Language in Science Fiction and Fantasy. The Question of Style*. London: Continuum.
- McCracken, S. 1998. *Pulp. Reading Popular Fiction*. Manchester: Manchester University Press.
- McEnery, T. and A. Hardie. 2012. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
- McIntyre, D. 2012. 'Corpus stylistics in the classroom' in M. Burke, S. Csabi, L. Week and J. Zerkowitz (eds) *Current Trends in Pedagogical Stylistics*, pp. 113-25. London: Continuum.
- Montoro, R. 2012. *Chick Lit: The Stylistics of Cappuccino Fiction*. London: Bloomsbury.
- Montoro, R. 2018 (fc). 'Investigating syntactic simplicity in popular fiction: a corpus stylistics approach' in R. Page, B. Busse and N. Nørgaard (eds) *'Only Connect': Language, Text and Context*. London and New York: Routledge.
- Moretti, F. 2010. 'History of the novel, theory of the novel', *Novel: A Forum on Fiction* 43 (1), pp 1-10.
- Nash, W. 1990. *Language in Popular Fiction*. London: Routledge.
- Oxford Text Archive, available at: <http://ota.ox.ac.uk/> [last accessed 10th April 2018]
- Pullum, G. 2004. 'The Dan Brown Code', *Language Log*, available at: <http://itre.cis.upenn.edu/~myl/language-log/archives/000844.html> [last accessed 9th March 2018]
- Ramet, C. 1999. *Ken Follett: The Transformation of a Writer*. Bowling Green, OH: Popular Press.
- Rayson, P. 2009. *Wmatrix: A Web-based Corpus Processing Environment*. Computing Department: Lancaster University, available at: <http://ucrel.lancs.ac.uk/wmatrix/> [last accessed 12th April 2018]
- Semino, E. and M. Short. 2004. *Corpus Stylistics: Speech, Writing and Thought Presentation in a Corpus of English Writing*. London: Routledge.
- Stockwell, P. 2000. *The Poetics of Science Fiction*. Harlow, Essex: Longman.

- Taylor, J. 1990. 'From pulpstyle to inner space: The stylistics of American New-Wave SF', *Style* 24 (4), pp. 611-627.
- Tucker, S. 2003. *The Last Year of Being Single*. London: Red Dress Ink.
- Werner, V. 2018. 'Linguistics and pop culture: setting the scene(s)' in V. Werner (ed.) *The Language of Pop Culture*, pp. 3-26. New York and Abingdon, Oxon: Routledge.
- Whelehan, I. 2002. *Helen Fielding's Bridget Jones's Diary. A Reader's Guide*. New York and London: Continuum.