

Corpus Stylistics  
and  
Henry James's Syntax

Lesley Moss

A thesis submitted for the degree of  
Doctor of Philosophy at UCL  
2014

Department of English Language and Literature  
UCL

I, Lesley Moss, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

.....

**Lesley Moss**

# Abstract

---

The starting point of this dissertation is a methodological question: how can corpus stylistics be used to analyse the syntax of literary fiction? A comparison of the syntax of Henry James's late style in *The Golden Bowl* (1904) and his early style in *Washington Square* (1881) was used as a case study. While James's late style is very widely discussed by literary critics and often seen as 'difficult', there has been very little evidence offered to substantiate this description.

Within the extensive field of Henry James studies, there have been few linguistic descriptions of James's prose. To remedy this, I compiled The *Henry James Parsed Corpus (HJPC)* from five chapters from each of the two novels.

My analysis of the corpus showed that *The Golden Bowl* is more syntactically complex than *Washington Square* in a number of ways but only in sentences which do not contain direct speech. James's idiosyncratic use of parenthesis was defined precisely using syntactic criteria and named *delay*. *The Golden Bowl* has more delay than *Washington Square* but also only in non-speech sentences. Only a small number of sentences have very high numbers of dependent clauses and/or delay. I argue that these exceptional sentences create the impression that the later text is homogeneously difficult. My research shows that this impression is deceptive; in fact the overwhelming majority of sentences in *The Golden Bowl* are no more syntactically complex than those of *Washington Square*.

A secondary use of the *HJPC* is to assist close reading. Chapter outlines of the central chapter of each novel were generated and were found to mirror plot developments and dialogue sections. Salient sentences highlighted many key moments in the plot, or revealed aspects of characters' personalities.

For my parents, Dorothy and Sam Moss,  
who taught me to love books and reading.

# Table of Contents

---

Abstract	3
Table of Figures	9
Table of Tables	11
Acknowledgements	12
List of Abbreviations	13
Chapter 1 Introduction	14
1.1 Introduction: the research question	14
1.2 Style and stylistics	14
1.2.1 Style	15
1.2.2 Stylistics	17
1.3 Corpus linguistics	25
1.4 My research process	28
1.5 The organisation of this dissertation	28
Chapter 2 Corpus Stylistics	30
2.1 Introduction	30
2.2 Defining corpus stylistics	31
2.3 Pre-computer studies	31
2.4 Early computer studies	33
2.5 The influence of corpus linguistics	36
2.6 Multidimensional Analysis	39
2.7 Authorship attribution	40
2.8 Speech, thought and writing	44
2.9 Lexical studies	45
2.9.1 Frequency	46
2.9.2 Collocation	48
2.10 Combining corpus stylistic methods	53
2.11 Corpus stylistics and syntax	57
2.12 Conclusion	62
Chapter 3 Henry James and Style	64

3.1	James's literary criticism .....	64
3.2	Literary criticism on James .....	67
3.2.1	Close reading of James .....	68
3.2.2	James and Modernism .....	70
3.2.3	Drama, picture and point of view .....	71
3.2.4	James's metaphor and imagery .....	73
3.2.5	James's language .....	75
3.2.6	Quantitative studies of James's style .....	77
3.3	Style in <i>Washington Square</i> and <i>The Golden Bowl</i> .....	88
3.3.1	<i>Washington Square</i> .....	88
3.3.2	<i>The Golden Bowl</i> .....	93
3.4	Conclusion .....	101
Chapter 4	Compiling the Corpus: Methodology .....	102
4.1	Sourcing the texts .....	102
4.2	ICECUP .....	104
4.3	The <i>Henry James Parsed Corpus</i> – ICECUP adapted .....	108
4.4	Specialised notation .....	113
4.4.1	Direct speech .....	113
4.4.2	Additional parsing units (APU) .....	114
4.4.3	Delay .....	117
4.4.4	Coordination .....	127
4.5	Hard cases .....	130
4.5.1	Self-correction .....	130
4.5.2	What is not delay? .....	130
4.5.3	Phrasal verbs .....	131
4.5.4	Apposition .....	132
4.6	Conclusion .....	134
Chapter 5	Analysing the Data .....	135
5.1	Vital statistics .....	136
5.2	Difficulty and complexity .....	138
5.2.1	Other measures of complexity .....	138

5.2.2	Coordination and dependency .....	140
5.2.3	Complexity and the representation of speech .....	146
5.2.4	Exceptional sentences .....	150
5.2.5	Distribution of dependency.....	152
5.3	Multiple main clauses.....	155
5.4	Delay .....	156
5.4.1	Delay and the representation of speech .....	159
5.4.2	Exceptional sentences .....	160
5.4.3	Distribution of delay scores .....	162
5.5	Conclusion.....	165
Chapter 6	Corpus Close Reading .....	169
6.1	Introduction .....	169
6.2	The two chapters .....	170
6.3	Chapter comparison.....	172
6.4	Sentence comparison.....	178
6.5	<i>Washington Square</i> Chapter 18.....	182
6.5.1	Profiles.....	182
6.5.2	Salient sentences.....	194
6.6	<i>The Golden Bowl</i> Chapter 22.....	199
6.6.1	Profiles.....	200
6.6.2	Salient sentences.....	213
6.7	Conclusion.....	218
Chapter 7	Conclusion.....	219
7.1	Introduction .....	219
7.2	Henry James's syntax .....	219
7.2.1	Critical opinions of James's style.....	219
7.2.2	Results of my analysis .....	221
7.2.3	An alternative approach.....	223
7.3	Corpus stylistics .....	224
7.4	Further research .....	225
7.4.1	Henry James's syntax.....	225

7.4.2 Corpus stylistics and syntax .....	226
7.5 Conclusion.....	227
Bibliography.....	228
Appendix 1 <i>Washington Square</i> Chapter 18.....	239
Appendix 2 <i>The Golden Bowl</i> Chapter 22 .....	244



# Table of Figures

---

Figure 3-1 James's three phases (Hoover, 2007, p. 182) .....	85
Figure 3-2 Chiastic inversion diagram .....	98
Figure 4-1 The sectors of a node (Nelson, et al., 2002b, p. 22) .....	104
Figure 4-2 An ICE-GB sentence displayed in an ICECUP tree .....	106
Figure 4-3 Part of WS03:7 illustrating phrase node use .....	109
Figure 4-4 GB05:3 before hand processing .....	111
Figure 4-5 GB05:3 after hand processing .....	112
Figure 4-6 GB22:67 showing markup .....	116
Figure 4-7 Structure of a kernel main clause .....	118
Figure 4-8 Structure of a kernel subordinate clause .....	118
Figure 4-9 Phrase structure tree .....	120
Figure 4-10 GB22:65 showing delay.....	122
Figure 4-11 GB22:79 showing detached function and delay.....	126
Figure 4-12 GB01:148 illustrating coordinated clauses and adjectives .....	128
Figure 4-13 GB01:148 detail showing coordination feature at phrase level .....	129
Figure 4-14 Part of WS05:125 illustrating apposition.....	133
Figure 5-1 Vital statistics compared .....	137
Figure 5-2 Part of WS01:12 illustrating clausal hierarchy .....	142
Figure 5-3 WS18:57 showing coordinated clauses.....	143
Figure 5-4 Part of GB22:9 illustrating dependent clauses .....	144
Figure 5-5 Coordinated and dependent clauses.....	145
Figure 5-6 Relative proportion of speech and non-speech sentences .....	147
Figure 5-7 Relative proportion of speech and non-speech clauses .....	147
Figure 5-8 Main and dependent clauses in speech sentences .....	148
Figure 5-9 Main and dependent clauses in non-speech .....	149
Figure 5-10 Distribution of dependent clauses .....	154
Figure 5-11 Multiple main clauses .....	156
Figure 5-12 Instances of delay per total words .....	157
Figure 5-13 Instances of delay in speech and non-speech .....	159
Figure 5-14 Distribution of delay scores.....	164
Figure 6-1 WS18 in context.....	173
Figure 6-2 Speech sentences in <i>WS</i> .....	174
Figure 6-3 Instances of delay per 1,000 words .....	174
Figure 6-4 Dependent clauses in speech and non-speech.....	175
Figure 6-5 GB22 in context .....	176
Figure 6-6 Speech sentences in <i>GB</i> .....	176
Figure 6-7 Instances of delay per 1,000 words .....	177

Figure 6-8 Dependent clauses in speech and non-speech.....	178
Figure 6-9 Comparison of coordination and complexity scores .....	182
Figure 6-10 WS18 Number of words per sentence.....	184
Figure 6-11 WS18 Coordination Score per Sentence.....	188
Figure 6-12 WS18 Complexity score by sentence.....	191
Figure 6-13 GB22 Number of words per sentence .....	201
Figure 6-14 GB22 Coordination score per sentence .....	204
Figure 6-15 GB22 Complexity score per sentence .....	207

# Table of Tables

---

Table 3-1 Image density in James's fiction	73
Table 3-2 Image density per 1,000 words by decade	74
Table 3-3 Chatman's figures - collated from different parts of <i>The Later Style of Henry James</i> (Chatman, 1972)	82
Table 5-1 Vital statistics of the <i>HJPC</i>	136
Table 5-2 Ratio of <i>WS</i> to <i>GB</i>	137
Table 5-3 Exceptional sentences	151
Table 5-4 Exceptional sentences (2)	160
Table 6-1 Comparative statistics	172
Table 6-2 WS18 distribution of long sentences	185
Table 6-3 WS18 distribution of complex sentences	189
Table 6-4 WS18 Speaker comparisons	193
Table 6-5 WS18 Salient sentences	195
Table 6-6 GB22 Longest sentences	202
Table 6-7 GB22 Distribution of long sentences	203
Table 6-8 GB22 Distribution of complex sentences	205
Table 6-9 GB22 Speaker comparisons	212
Table 6-10 Salient sentences	213

# Acknowledgements

---

Many people have supported me during this long project. I would like to thank Dr Peter Wilson, who first introduced me to stylistics, Professor Bas Aarts, who taught me that syntax is great fun and then generously took on my PhD supervision in my second year, Professor Philip Horne, my second supervisor, for his encouragement and help, and my original supervisor, Professor Gerald Nelson, who suggested Henry James as a subject.

Many members of the Survey of English Usage at UCL have helped me along the way. I would like to thank Yordanka Kavalova and Dr Gabriel Ozon, who taught me how to parse in ICECUP, Dr Kathryn Allen, Dr Joanne Close, Dr Jillian Bowie, Dan Clayton and Seth Mehl, who all patiently discussed Henry James's syntax from time to time, and Isaac Hallegua for all the gluten-free cakes.

I would also like to thank Stephen Kirk, who encouraged me to start the PhD, Radha Chakraborty, who has proofread, discussed statistics and theories and generally encouraged me. I am also grateful to my wonderful sons, Daniel and Simon Kirk, who have listened to me through all the ups and downs, and my beloved husband, Adrian Shindler, an invaluable reference librarian, proof reader and patient PhD-widower.

Finally, I would particularly like to thank Sean Wallis, who spent much time and effort teaching me how to compile a corpus in ICECUP and building the initial framework for me, helping me to manage my corpus, and patiently trying to explain statistics to me.

# List of Abbreviations

---

A	adverbial	intr	intransitive
ADJ	adjective	main	main clause
ADV	adverb	montr	monotransitive
AJHD	adjective phrase head	MVB	main verb
AJP	adjective phrase	N	noun
appos	appositive	neg	negative
APU	additional parsing unit	NONCL	non-clause
ART	article	NP	noun phrase
attru	unmarked attributive	NPHD	noun phrase head
AUX	auxiliary	NPPO	noun_phrase postmodifier
AVB	auxiliary verb	NPPR	noun phrase premodifier
AVHD	adverb phrase head	OD	direct object
CJ	conjoin	OP	operator
CL	clause	oquo	open quotes
com	common	P	prepositional
comment	comment clause	PARA	parataxis
comp	comparative	pass	passive
CONJUNC	conjunction	PC	prepositional complement
CONNEC	connective	per	period
COOR	coordinator	pers	personal
coord	coordinating	phras	phrasal
coordn	coordination	plu	plural
cop	copulative	poss	possessive
cquo	close quotes	PP	prepositional phrase
CS	subject complement	PREP	preposition
DEFUNC	detached function	pres	present
depend	dependent	PRON	pronoun
DISMK	discourse marker	prop	proper
DT	determiner	PRTCL	particle
DTCE	central determiner	PU	parsing unit
DTP	determiner phrase	PUNC	punctuation
edp	-ed participle	rel	relative
encl	enclitic	sing	singular
GB	that part of the <i>HJPC</i> compiled from <i>The Golden Bowl</i>	S	sentence
ge	general	SBHD	subordinator_phrase head
HJPC	<i>Henry James Parsed Corpus</i>	SU	subject
ICECUP	International Corpus of English Corpus Utility Program	V	verb
indef	indefinite	VB	verbal
infin	infinitive	VP	verb phrase
inten	intensifying	WS	that part of the <i>HJPC</i> compiled from <i>Washington Square</i>
		zsub	zero subordinate

Sentences are indicated by their book, chapter and number, e.g. sentence 12 of chapter 2 in *Washington Square* is abbreviated to WS02:12, sentence 3 of chapter 22 of *The Golden Bowl* is abbreviated to GB22:03.

# Chapter 1 Introduction

---

## 1.1 Introduction: the research question

The research question which motivates this project is a methodological one: how can the methods of corpus stylistics be used to explore the syntax of literary texts? I aim to elucidate how useful such an approach is in aiding literary criticism, and what difficulties arise. Methodological research requires subject matter and so a second research question is needed. Here the question is what are the differences, if any, between Henry James's early and late novels in terms of syntax?

This project fills a gap in current corpus stylistics. Although this sub-discipline is growing within the wider context of stylistics generally, it is focused almost exclusively on lexis and not on syntax. In particular, the creation of a fully parsed corpus from literary texts is almost unknown, so this project affords new opportunities for analysis.

The need for the project in the context of the case study is that Henry James is very widely described as having a 'difficult late style'; Chatman's starting point in his analysis of James's late style is 'the usual characterization of James' later style as 'difficult' (1972, p. 2). He goes on to interpret this difficulty as being due to 'abstractness'. However, there is very little other work on what the late style actually consists of in terms of James's language, and there is a particular gap both in contrasting the late and early styles, and in analysis of James's syntax. A further omission is a clear differentiation between the subjective cognitive response of considering a novel 'difficult' and the rigorous examination, focussed on the text of a novel, which stylistics entails. My goal is to bring objective, quantitative data into the study of James's style and into the study of literary style generally, while acknowledging that a subjective element and critical judgement can never, and should never, be eliminated.

In this introductory chapter the two disciplines which have contributed to corpus stylistics are examined – style and stylistics in section 1.2 and corpus linguistics in section 1.3.

## 1.2 Style and stylistics

'Style' is a contested term with a relatively vague meaning in common speech, a number of competing technical definitions and an implied association with the academic field of stylistics. In this context, it is perhaps inevitable that stylistics, the study of style, has also attracted competing arguments and criticisms. The debate around the term 'style' is outlined in section 1.2.1 while discussions on the role and shortcomings of stylistics are described in section 1.2.2.

### 1.2.1 Style

In 1964 Ohmann saw the concept of style as extremely vague, even when implicitly defined narrowly as literary style:

A style is a way of writing – that is what the word means. And that is almost as much as one can say with assurance on the subject, which has been remarkably unencumbered by theoretical insights. (Ohmann, 1964, p. 423)

He asserts that a reader has an intuitive feeling for the style of a text they read, separating that style from its content – '[a] feeling for the quiddity of a writer's linguistic method, a sense of differences between stretches of literary discourse which are not differences in content' (Ohmann, 1964, p. 423). David Lodge (1966) cites the clarification of the concept of style as one of the tasks of modern stylistics as well as developing 'more precise, inclusive, and objective methods of describing style than the impressionistic generalizations of modern criticism' (Lodge, 1966, p. 55).

Consideration of what literary style consists of and how it might be studied is often traced back to ancient times. In more recent years, an important stage in its analysis was a 1958 conference held at Indiana University in which academics from a number of disciplines came together to discuss the concept and how it might be studied in a more scientific way. Speakers represented psychology, linguistics and literary criticism.

In our conference ... a deliberate and self-conscious attempt was made to initiate a departure from the perpetual humanistic engagement in the solution of a subtle and elusive puzzle - the fluid and dissonant notion of style - by offering an opportunity for experts in philosophical speculation to commingle (if not outrightly collaborate) with men of scientific temperament. (Sebeok, 1960, pp. 4-5)

Discussion centred on the question of how to identify an individual's style, how a writer makes linguistic choices, and whether poetic language should be regarded as deviant from some assumed norm. The conference participants grappled with the problem of form and content, and to what extent these are separate or separable. One problem of these papers is the different definitions of style offered by each speaker, both within disciplines and between them. For example, Osgood, a psychologist, focused on measurable features of style, treating the phenomenon as one of deviation.

Style is defined as an individual's deviations from norms for the situations in which he is encoding, these deviations being in the statistical properties of those structural features for which there exists some degree of choice in his code. (Osgood, 1960, p. 293)

The 1958 conference also included Jakobson's much-cited paper 'Linguistics and Poetics' dealing with poetic language as choice and including the various functions of poetic language (Jakobson, 1960).

The issues identified in 1958 remain current. Style remains a vague term. Wales gives five different definitions but prefaces them by explaining that 'although style is invoked very frequently in literary criticism, translation studies, sociolinguistics and especially stylistics it is very difficult to define' (Wales, 2011, p. 397). The question of to what extent form and content should be regarded as separate, or can be analysed separately, continues to be debated.

At a second conference in Italy in 1969 Barthes (1971) discussed the usefulness of a binary concept of style and content, tracing the history of such a division to classical rhetoric. He also discussed the idea of style as a deviation from a norm, and the problem of then deciding what the norm is. He identified two stylistic values – simplicity and strikingness – and suggested there was a need to describe literary language as a whole before beginning to discuss the style of a particular author. Some characteristics of literary style which he identified are inversions, order of complements, archaisms and metaphors, and he emphasised the layered quality of literary language (a term he prefers to 'style').

Carter discusses the concept of literary language in an encyclopaedia entry on the topic (1994), identifying it as a 20<sup>th</sup> century idea which Russian and Czech formalists were among the first to define. They proposed that literary language was characterised by de-automatization, using devices like foregrounding.<sup>1</sup> However, this presupposes a language norm from which a novelist was deviating and, again, this norm was hard to define. Carter quotes Widdowson's description:

[The] language of a literary work should be fashioned into patterns over and above those required by the actual language system...the effect of this patterning is to create acts of communication which are self-contained units, independent of a social context and expressive of a reality other than that which is sanctioned by convention. (Carter, 1994, p. 2248)

Carter proposes the study of the extent of literariness in a particular language text. This would be a continuum and a number of tests could be applied to a text to quantify its literariness, such as:

- density of polysemic effects;
- the displaced character of the interaction between author and reader;
- the extent of independence of the text from other media (such as pictures, diagrams);
- the kinds of re-registrations of words and phrases from one context to another;
- and the density of iconic or representational uses of language (Carter 1994 p.2249)

---

<sup>1</sup> Foregrounding is described in more detail in section 1.2.2



Wales concurs with this view; 'stylisticians today tend to agree with literary critics in seeing a continuum of literariness in language across non-literary as well as literary discourses (whether high or noncanonical) (Wales, 2006, p. 215).

Traugott and Pratt, in an excerpt collected by Carter and Stockwell (2008), discuss the concept of style as authorial choice. Their concept is developed in the context of the application of generative grammar to literary style, a practice which is now seldom pursued. They see content as conceptually separate to the form of a text, which will be conditioned by the pragmatic circumstances in which it is used, for instance to address a child. They identify choice as applicable to both content and form and, while they agree that the choices an author makes may be seen as a deviation from a grammatical norm, this is only one of the available choices (and any norm is not fixed but becomes established by the text itself). This view is echoed by Verdonk:

In this view, style is seen as the making of conscious and unconscious choices of certain linguistic forms and structures in preference to others that could have been chosen but were not. (Verdonk, 2006, p. 203)

These choices may be at any level of the text from graphology and phonology through lexis and syntax to semantics and pragmatics. Verdonk sees the semantic content and style of a text as essentially unified but separated for the purpose of analysis.

In the context of this amount of disagreement about what literary style is, it is unsurprising that stylistics is also a contested discipline, both as to its methods and its academic standing.

### **1.2.2 Stylistics**

Stylistics emerged as a field from a number of different approaches to literary criticism as well as the application of linguistics to literary texts. Wales (2006) describes influences from Russia and Eastern Europe, France and Germany as well as the UK and the United States. Both Wales, and Carter and Stockwell (2008), date the establishment of stylistics as a truly separate field from the 1960's, in Wales's case explaining this by developments in linguistic theory, particularly Halliday's functional grammar. Carter and Stockwell point out that earlier approaches (such as close reading, particularly in the UK, and its development into New Criticism in the US) already focused on the text of literary works but the addition of linguistic insights led to 'the different influences [being] integrated into a set of conventions for analysis' (Carter & Stockwell, 2008, p. 293). In 1966 Lodge saw Spitzer as 'the father of "the New Stylistics' (Lodge, 1966, p. 56), citing his concern to relate linguistic structures in literary texts to aesthetic effects (and even the psychology of the author). While Lodge rejects the psychological aspects of Spitzer's work, he approves of his emphasis on taking an intuitive response to a

text as a starting point for a more analytical exploration, which may confirm or negate the original reaction. In this sense Lodge sees stylistics as a more objective approach which needs to be added to the aesthetic and evaluative point of view of the literary critic.

In his introduction to the papers of the 1969 conference on stylistics, Chatman (Chatman, 1971) identifies five main strands in 'modern literary stylistics':

- Russian-Formalist-Jakobsonian school
- French *structuralistes* influenced by Russian and Czech theory
- British school influenced by Firth
- American school influenced by New Criticism and Bloomfield
- Continental school influenced by Spitzer (Chatman, 1971, p. ix).

At the conference Josephine Miles (1971) considered the problem of how far to study individual style, and said that stylistics would be mainly comparative, between particular texts, authors or genres. She highlighted the usefulness of looking at co-occurrence, which is a current interest in corpus stylistics (see section 2.9.2).

The ideas of early Russian and Czech linguists remain influential through the concept of *foregrounding*. The term 'foregrounding' is a translation by P. L. Garvin (1964) of the term *aktualisace* used by the Prague School in the 1930's. Garvin explains that the term 'refers to a stimulus not culturally expected in a social situation and hence capable of provoking special attention' (Garvin, 1964, p. viii).

Wales identifies members of the School like Mukařovsky and Havránek, who followed the earlier Russian Formalists in believing that:

it was the function of poetic language to surprise the reader with a fresh and dynamic awareness of its linguistic medium, to de-automatize what was normally taken for granted, to exploit language aesthetically. (Wales, 2011, p. 166)

Leech (2008) explains that such a surprising stimulus slows the processing of the reader of a text (or viewer of a visual art object) so that they pay it more attention and give more consideration to their interpretation. This process of showing something in a new way is also known as 'defamiliarization' (Leech, 2008, p. 4). The surprising element in the stimulus is that it is a departure from a norm.

Formally, foregrounding is a deviation, or departure, from what is expected in the linguistic code or the social code expressed through language; functionally, it is a special effect or significance conveyed by that departure. (Leech, 2008, p. 3)

Short emphasises the cognitive nature of foregrounding.

Deviation, which is a *linguistic* phenomenon, has an important *psychological* effect on readers (and hearers). If a part of a poem is deviant, it becomes especially noticeable, or perceptually prominent. We call this psychological

effect **foregrounding**. (Short, 1996, p. 11) [*Italics and bold style Short's own.*]

(Foregrounding has particularly been applied to the stylistic analysis of poetry but as Jeffries and McIntyre point out, foregrounding devices can be found in other types of literary and non-literary texts as well (Jeffries & McIntyre, 2010, p. 2).)

Wales describes foregrounding as being 'achieved by a variety of means, which have been largely grouped under two main types: deviation and repetition' (Wales, 2011, p. 167). However, all foregrounding involves a deviation from the expected linguistic norm at some level of language. Levin considers foregrounding to be key to the language of poetry, allowing a poem to draw attention to itself as an object separate to the meaning of its words. 'It can be shown that most, if not indeed all, of poetry's characteristic devices exemplify deviation in one way or another' (Levin, 1965, p. 225). Levin identifies two types of deviation:

Internal - 'that type of deviation which takes place against the background of the poem, where the norm is the remainder of the poem in which the deviation occurs'.

External - 'that type where the deviation is to be explicated against some norm which lies outside the limits of the poem in which the deviation occurs'. (Levin, 1965, p. 226)

However Leech (2008) describes three kinds of deviation, renaming internal deviation as tertiary deviation, and dividing external deviation into that which differs from the norms of language generally, and that which deviates both from general language norms and also from accepted expectations for a literary text.

The second method of foregrounding, repetition or parallelism, involves repetition at any level of language beyond what is taken to be usual according to an internal or external norm. Short points out that the reader is likely to infer a link between parallel items, such as alliterative or assonant words.

These concepts have been widely applied in stylistics from the early twentieth century to the present day, and so have also been used in corpus stylistics. Research has been done directly on the theory that foregrounding devices slow the reading process (da Costa Fialho, 2007), on the extension of foregrounding theory to novels from poems and short texts (Sopčák, 2007), on the relationship between foregrounding and aesthetic appreciation (van Peer, et al., 2007), and on the effects of foregrounding on authorial innovation (Martindale, 2007). The theory also has relevance for this research project, as will be explained in Chapter 2 and further discussed in Chapters 5, 6 and 7.

Carter and Stockwell cite Leech's 1969 *A Linguistic Guide to English Poetry* as 'a landmark of early stylistics...demonstrating the familiar double-advantage of stylistics as a contribution to linguistic theory and to literary criticism simultaneously' (Carter & Stockwell, 2008, p. 60). In that book, Leech defines

stylistics as 'the study of the use of language in literature' (Leech, 1969, p. 1), and the area where literary criticism and linguistic analysis overlap. However, nearly 40 years later, in 2008, Leech points out that the word 'stylistics' is still not fully accepted, with other terms proposed to define a field which he calls an 'interdiscipline' (Leech, 2008, p. 1). Leech identifies a number of causes for this uncertainty which surrounds stylistics: innovative developments in linguistics, the discipline from which stylistics draws much of its methodology, as well as uncertainty about the canonical status of literature and about the aims of literary criticism. Nevertheless, Leech espouses the term 'stylistics' to describe 'the field acting as a bridge between linguistic and literary studies' (Leech, 2008, p. 2).

Birch (1994), in an encyclopaedia article, defines stylistics quite narrowly as a practical, analytic exercise. Stylistics has had to fight to justify its place by 'displaying a facility with linguistic description which contributes substantially to the critical vocabulary of textual analysis' (Birch, 1994, p. 4380). There remains a problem with how far evaluation and appreciation are derived from stylistic analysis, and it remains true that in deciding what to analyse, the intuitive focus of the critic is often the starting point. Widdowson (2008), writing in 1972, also emphasises the basic difference of focus between linguistics and literary critics; where the linguist may see literary language as providing evidence for a grammatical theory or as an example of a deviation from a linguistic norm, for the critic a literary text is a unique object with an aesthetic value. Widdowson espouses a search for common interests between the two disciplines in what can be seen as a goal for the field of stylistics. However, he is also cautious about the wider application of linguistic analysis; 'after the linguist has finished his analysis, there remains the major task of establishing its significance. We should not be surprised if its significance turns out to be marginal' (Widdowson, 2008, p. 30).

This tension between a desire for objectivity and criticism that stylistics lacks an adequate methodology is a continuing theme in descriptions of the field. In recent times, stylisticians have argued that both purposes must be fulfilled. For example, Cureton, in an encyclopaedia article from 2003, emphasises the importance of stylistics and argues that it includes both a measure of objectivity and an aesthetic evaluation.

We respond aesthetically to language when our dominant response is to appreciate some quality of the language, independent of other ends to which that language is directed. The problem then is to identify some of these qualities. (Cureton, 2003, p. 467)

Cureton identifies a number of different approaches within the field of stylistics. He contrasts a view of aesthetic language as deviant in comparison to 'everyday' language with a focus within the text itself without outward comparison. Cureton includes syntax as one of the areas which can be explored in a stylistic analysis,

suggesting that the syntactic effects noted will be conditioned by the grammatical theory employed. However, Cureton underlines his view of the embattled status of stylistics, although not without hope that it might improve.

At present, relations between stylistics and its neighboring disciplines are tentative at best. To most linguists, stylistics is a peripheral sort of applied analysis; to most literary critics, it is a laborious dwelling on irrelevancies. (Cureton, 2003, pp. 470-471)

One of the most cited stylistic analyses is Halliday's paper on the language of William Golding's *The Inheritors*, which was given at the 1969 conference on stylistics discussed above. Halliday locates his analysis of Golding's novel within his more general theory of the function of language. While identifying three basic functions of language, he explains that it is particularly the 'textual function' which has relevance for stylistics. 'It is through this function that language makes links with itself and with the situation; and discourse becomes possible, because the speaker or writer can produce a text and the listener or reader can recognize one' (Halliday, 1971, p. 334). Halliday links the concept of foregrounding in a literary text to what he calls 'prominence' or linguistic highlighting. 'Foregrounding, as I understand it, is prominence that is motivated' (Halliday, 1971, p. 339). He prefers to avoid defining prominence as a deviation from a norm because of the difficulties of choosing a suitable norm and because different readers are likely to have different norms in mind. This does not deter him from stating that it will be possible to identify prominence statistically by comparison with the language as a whole.

Halliday tackles and dismisses two criticisms of stylistic analysis. He asserts that accusations that individual style is too idiosyncratic to be analysed quantitatively miss the point.

If there is such a thing as a recognizable style, whether of a work, an author, or an entire period or literary tradition, its distinctive quality can in the last analysis be stated in terms of relative frequencies, although the linguistic features that show significant variation may be simple and obvious or extremely subtle and complex. (Halliday, 1971, p. 343)

Halliday also rejects the idea that quantitative qualities of a text cannot be relevant to style because they are not detected by the reader. For him it seems likely that the reader does notice these frequencies, even if subconsciously, but in any case he is interested in author choice rather than reader response. 'If in the selections [the writer] has made there is an unexpected pattern of frequency distributions, and this turns out to be motivated, it seems pointless to argue that such a phenomenon could not possibly be significant' (Halliday, 1971, p. 344). The difficulty which remains is to ascertain whether what is prominent is also stylistically relevant, particularly as the degree to which a feature is stylistically relevant may not correspond precisely to its degree of prominence. It will be necessary for the stylistician to move beyond the analytic data to make a stylistic appraisal. At this

point, Halliday returns to the judgement of the critic as the final arbiter of what is stylistically important, even if informed by quantitative data they have found in the text. He paraphrases Ullmann as saying 'in stylistics we have *both* to count things *and* to look at them, one by one' (Halliday's italics) (Halliday, 1971, p. 347).

To illustrate his thesis, Halliday analyses three passages from Golding's *The Inheritors*, mainly focusing on the syntax. He identifies a Neanderthal language, a Homo sapiens language and an intermediate language with reference to the frequency of different syntactic features such as the use of intransitives or lack of noun modifiers. Each language relates to the plot and theme of the novel at that stage; for example, Halliday suggests that the predominance of inanimate agents in the Neanderthal language highlights their powerlessness in the face of the changes in their world. The frequency of inanimate agents is a quantitative fact which has been chosen by Golding (consciously or unconsciously) and is therefore foregrounded and has stylistic impact.

In responding to questions at the conference at which this paper was first presented, Halliday described his analytic process. This began with an intuitive recognition of the prominence of an item which was then counted. If the prominence relates to a theme (which has also been noticed intuitively), then it is considered to be stylistically relevant.

This seminal paper by Halliday was one of the targets of Fish's critical paper on stylistics delivered at a session of the English Institute chaired by Seymour Chatman in 1973. Fish characterises stylistics as a reaction to the current literary criticism.

Stylistics was born of a reaction to the subjectivity and imprecision of literary studies. For the appreciative raptures of the impressionistic critic, stylisticians purport to substitute precise and rigorous linguistic descriptions and to proceed from these descriptions to interpretations for which they can claim a measure of objectivity. Stylistics, in short, is an attempt to put criticism on a scientific basis. (Fish, 1980, pp. 69-70)

Fish objects to the goal of objectivity, which he considers impossible. He reviews work by Milic, Ohmann and Thorne, pointing out conclusions which he considers tenuous and arguments which he describes as circular. He particularly wants to assert that linguistic structures in a text should not be linked to a particular meaning; in a different text, the same structure might have a different meaning.

Fish has a number of criticisms of Halliday's paper, ranging from the complexity of his grammatical theory to a confusion between syntactic terminology and descriptive categories. Fish rejects Halliday's interpretation of the effects of Golding's unusual syntax both on the immediate impact of the passage and its wider implications. A more cogent criticism is Fish's objection that 'Halliday's interpretation precedes his gathering and evaluating of the data, and it, rather than any ability of the syntax to embody a conceptual orientation, is responsible for the

way in which the data are read' (Fish, 1980, p. 82). However, Halliday himself describes his process as beginning from an intuitive reading, which guides his search for data.

Fish's most fundamental objection is to Halliday's focus on the writer and the text rather than the reader and their interpretation of the text. Without reference to a reader's interpretation, for Fish a text is 'quite literally meaningless' (Fish, 1980, p. 84).

I do not ... deny that the formal distinctions Halliday uncovers are meaningful; but where he assumes that they *possess* meaning (as a consequence of a built-in relationship between formal features and cognitive capacities), I would argue that they *acquire* it, and that they acquire it by virtue of their position in a structure of experience. (Fish, 1980, p. 92)

Fish does not completely reject stylistics but argues for an 'affective stylistics' in which the constituents of the text are understood as interpreted by a reader with their particular preconceptions and context. 'Interpretive acts are being described' (Fish, 1980, p. 93).

However, Carter (2003) does not accept Fish's criticism. For him, an informed account of the language of a literary work is a vital part of transforming a reader's intuitive response to a text into an account which is supported by the features of that text. He does not wish to argue that a definitive account will be formulated by a stylistic analysis but does claim that it will form the basis for a debate on the interpretation of the text.

Toolan considers Fish's criticism of stylistics in some detail, in the context of a wider consideration of the status and goals of the discipline. He places his argument in the context of a dispute over the meaning of the term 'style' and whether it should be the object of literary or linguistic study. Toolan sees literary style as on a continuum with the style of everyday language, but with a closer relationship of content and form and with a less immediate communicative function. Toolan has a nuanced, and to me convincing, view of the role of stylistics as an analytic method.

I propose that stylistics be viewed as a way rather than a method - a confessedly partial or oriented act of intervention, a reading which is strategic, as all readings necessarily are. The attraction of the "way" of stylistics lies in its attempt at public-ness, even as it acknowledges private-ness, unpredictability. If that is an unhappy compromise, it is not easy ... to conceive of a happier one or a less compromised happiness. (Toolan, 1990, p. 11)

This definition avoids Fish's accusation of a prescriptive reading from language to meaning, and of a blindness to the reader and their interpretation. On the contrary, Toolan emphasises that any reading is provisional and rooted in the cultural and personal context of the reader as well as their theoretical assumptions about language, explicit or implicit. At the same time, he asserts the usefulness of 'close

study of the language features of a text' (Toolan, 1990, p. 14) as part of a broader literary criticism. Linguistic features may convey meaning to a particular reader and there may be some consensus on these meanings within an interpretive community. So Toolan sees stylistics as 'a loose confederation of interpretative strategies in pursuit of a strategic interpretation of a text' (Toolan, 1990, p. 24), explicitly subjective but with the method of analysis open to inspection and discussion. Stylistics also provides the language with which this discussion can be pursued.

Many stylisticians reference Fish in their work, answering his criticisms of the field. Hoover responds more specifically in his book *Language and Style in The Inheritors*, where he re-analyses the language of *The Inheritors* using corpus stylistics. This analysis is described in section 2.11. For Hoover, there is no inherent problem with Halliday's idea that the syntax of a text conveys meaning. 'It seems unreasonable to deny that transitivity patterns can have definable and at least partly predictable semantic effects on readers and their responses' (Hoover, 1999, p. 22). Hoover is not claiming that linguistic analysis can provide a full interpretation of a text but it can show what is foregrounded in the language of a novel. For Hoover, the role of the reader in interpretation can be exaggerated.

[Linguistic features] must and do have semantic effects, just as they do in everyday language use. These syntactic features and the semantic effects they produce are part of the language itself, part of the system; they are not contributed by a particular reader, except in the rather trivial sense that he or she happens to be the reader in whom the effects are produced. (Hoover, 1999, p. 23)

Carter reprises the theme of the contested nature of stylistics in his essay with Stockwell (2008) which concludes their collection of articles on language and literature. They see stylistics as a discipline which has suffered from forming part of a number of academic disciplines and as 'sitting ... uncomfortably on the bridge between the linguistic and the literary' (Carter & Stockwell, 2008, p. 291). Literary critics have seen stylistics as 'too mechanistic and reductive' while linguists consider that it introduces non-measurable aesthetic evaluation. Despite this interdisciplinary position, Carter and Stockwell see stylistics as not only flourishing but also essentially coherent. Indeed they describe it as 'naturally the central discipline of literary study, against which all other current approaches are partial or interdisciplinary' (Carter & Stockwell, 2008, p. 292). Wales (2006) sees the multiple nature of stylistics more positively, characterising the discipline as one which integrates different linguistic and critical approaches in a constant evolution. She cites Leech and Short's work on the language of speech and writing, and the recent interest in cognitive stylistics as examples of this process. Nevertheless, like Carter and Stockwell, she sees stylistics as a marginalised field in academia, caught between literary and linguistics departments.



Jeffries and McIntyre (2010) defend the field of stylistics strongly, seeing it as an essential tool to clarify and explain the non-linguistic analysis of literary critics. They see objectivity as one of the defining characteristics of stylistics but note that this involves a willingness to open analysis to challenge and to change conclusions in the face of conflicting evidence. Such evidence must include the context as well as data from within the text. The form of the text is 'only the *beginning* of literary interpretation, which depends on the conjunction of form with particular content and also with particular contexts of production and reception' (Jeffries & McIntyre, 2010, p. 70) [Jeffries and McIntyre's italics]. While not accepting Fish's description, or rejection, of stylistics, the role of the reader must be included in the contextual factors to be considered, with the contingent interpretations which that implies. For Jeffries and McIntyre, this does not negate the existence of some textual qualities which are objectively analysable.

It is important to note that there remain some features of the text which *can* be identified and described irrespective of their intended and/or received effect. The interpretation of how such features may affect the writer's meaning and the reader's meaning is one of the more subjective aspects of stylistic analysis, though the link to textual features does at least achieve the scientific standard of explicitness that enables others to see how an interpretation is arrived at. (Jeffries & McIntyre, 2010, p. 71) [Their italics]

Jeffries and McIntyre locate stylistics in an academic context in which it has an essential role to play.

It is our contention that literary criticism as a discipline is stagnating because the unfalsifiability of the claims generated by this subjective approach, and the lack of focus on the text, makes critical discussion impossible. It is also our contention that stylistics offers a way out of this impasse. (Jeffries & McIntyre, 2010, p. 192)

### 1.3 Corpus linguistics

Corpus linguistics has been defined as

the empirical study of language relying on computer-assisted techniques to analyse large, principled databases of naturally occurring language. (Conrad, 2000, p. 548)

This also provides a definition of a corpus, while leaving unspecified how large it has to be, how it needs to be 'principled' and what 'naturally occurring language' is. A very simple definition of a corpus is 'a collection of texts (a "body" of language) stored in an electronic database' (Baker, et al., 2006, p. 48) but more generally a corpus is a large collection of texts which is machine-readable so that the corpus can be used for research. Baker, et al. distinguish a corpus from an archive by specifying that a corpus often contains data which has been chosen to be representative of a particular language at a particular period, or a genre; its defining feature is that it can then be used to provide normative data. Corpora are

also usually annotated in some way to provide information about the source of the texts and/or the language samples the texts include; for example, many corpora are annotated with the parts of speech of the words they contain. More recently McEnery and Hardie have provided a definition of corpus linguistics with a more instrumental focus; for them, corpus linguistics is a field which deals 'with some set of machine-readable texts which is deemed an appropriate basis on which to study a specific set of research questions' (McEnery & Hardie, 2012, p. 1). They also explain that a text does not have to be a written piece of language; the term includes speech samples, for example.

Despite these definitions, research which can be recognised as using corpus linguistic methods existed before computerisation. Some early corpus stylistics is described in section 2.2. Svartvik (2007) points out that, among others, the lexicographers of the *Oxford English Dictionary*, dialecticians and historical linguists have recorded natural language data on paper for many years. In English corpus linguistics, pioneering work was done at the Survey of English Usage at UCL. Starting in 1959 under the direction of Randolph (now Lord) Quirk, examples of the English of the time, both spoken and written, were initially collected on file cards before being computerised. The careful selection of an equal number of texts in a wide variety of genres fulfilled the requirement of a corpus which can be considered representative of educated British English. Over a million words were collected over 30 years, and in later years annotation of the data, specifically parsing the texts, and developing software to allow sophisticated searching completed the development of a modern corpus. Almost simultaneously, around a million words of American English were collected for the Brown Corpus, developed by Kucera and Francis at Brown University in Rhode Island. After these pioneering efforts, much larger computerised corpora of English were developed at a number of different academic institutions.

McEnery and Hardie distinguish two main types of corpus; the first is a monitor corpus, such as the Bank of English based at the University of Birmingham, initially under the supervision of John Sinclair. This type of corpus continuously collects new data. At the time of writing, the Bank of English contains 650 million words 'from a carefully chosen selection of sources, to give a balanced and accurate reflection of English as it is used every day' (Harper Collins). This is dwarfed by the 2.5 billion word Collins corpus which is added to every month to identify neologisms. The second type of corpus is a sample corpus; this consists of samples from a particular set of data. McEnery and Hardie include the corpus held at the Survey of English Usage in this class. The corpus developed for this project – the *Henry James Parsed Corpus* – is of the same type. This binary division of corpora does not accommodate those compiled from a complete set of specialist resources, such as those which comprise the total content of the journals *The Economist* and

*The Wall Street Journal*. In practice, despite many offered definitions, it is not clear what constitutes a corpus and whether it can simply be a collection of texts or whether, as most definitions assert, it must include some design or sampling criteria in its compilation.

Renouf distinguishes five phases in the development of corpora. From the 1960's onwards, 'small corpora' (Renouf, 2007, p. 28), which held one million words or less, were compiled in a standardised format of samples. By the 1980's similar types of corpora held many millions of words, while from the 1990's corpora were being added to in an open-ended manner, allowing for diachronic research. However, from 1998 Renouf suggests that the Web itself had become a corpus and that the future holds a next-level internet organisation with data and processing happening cooperatively.

The initial burst of development of corpora faced opposition from Chomsky and others who agreed with his view that language should be investigated using the intuitive introspection of the linguist rather than external data. Chomsky's approach is rationalist rather than empiricist and centres on linguistic competence rather than linguistic performance. In his opinion, corpora would not be useful for the exploration of language competence and could never represent the whole of a grammar. The performance data included in a large natural language corpus will not be entirely grammatically correct but will include errors and word play. Moreover, data is likely to be included in such a corpus because it is more frequent in the language represented, or purely by chance. In an interview with Bas Aarts, Chomsky compared corpus linguistics to the data collection and cataloguing of pre-scientific days (Aarts, 1999). Aarts responds by describing corpus linguistics as a methodology which may be employed for different theoretical approaches in linguistics. However, he also calls for the avoidance of mere reporting of data and calls for a 'focus on qualitative research' (Aarts, 1999, p. 8).

Since around the end of the 1980's, this controversy has been set aside by many linguists. McEnery and Hardie consider that 'corpus linguistics has become an indispensable component of the methodological toolbox throughout linguistics' (McEnery & Hardie, 2012, p. 226). Modern corpus linguistics encompasses a wide range of linguistic research, both in compiling corpora and in using their resources. For example, in Facchinetti's collection of papers (2007) both syntax and semantics are discussed, there are papers on both spoken and written English, and also research on the teaching of English as a second language and comparisons of the syntax of English and other languages.

As stylistics has always been influenced by linguistics, it was natural and inevitable that the development of corpus linguistics would also impact stylistics. Large corpora of English became a resource for stylisticians and, as seen in this project, the possibility of making literary texts themselves into corpora has proven

a fertile area for stylisticians. The details of the process in this project are described in section 1.4.

## 1.4 My research process

In this project I analyse and discuss the syntax of *Washington Square* and *The Golden Bowl*. *Washington Square* was published in 1881 and therefore dates from the end of James's early period of novel writing. *The Golden Bowl* was the last novel James completed and was published in 1904. I have taken these two novels as examples of James's early and late styles. (For further details of the selection of these novels, see Section 4.1)

Using five chapters from each novel, a corpus was compiled and named the *Henry James Parsed Corpus (HJPC)*. The two parts of the corpus, compiled from each novel respectively, are referred to as *WS* and *GB*, with the books themselves being referred to in full. A computer program called ICECUP was used to compile the corpus, which enabled it to be searched to characterise James's syntax within the two samples and, where a statistically significant result was obtained, within the two novels. A full description of the method of compiling and analysing the corpus is given in Chapter 4.

Once the *HJPC* was compiled, it was searched to identify the characteristics of the two books. 'Difficulty' was interpreted as partly consisting of syntactic complexity, which can be quantified. Other idiosyncratic features of James's style were identified, as well as the differences between the syntax used for dialogue and for non-speech. These processes and their results are described in Chapter 5. Statistical tests were employed to ascertain whether the results of searches of the *HJPC* were applicable to *Washington Square* and *The Golden Bowl* as a whole. This was achieved with the invaluable help of spreadsheets provided by Sean Wallis; for more details of this process see section 5.1.

A second application of the corpus data was to assist in a close reading of one chapter from each book, described in Chapter 6. Here quantitative data from the *HJPC* informs a more qualitative exploration of the plot, characterisation and language of the chapter.

## 1.5 The organisation of this dissertation

Following my discussion of the fields of stylistics and corpus linguistics, Chapter 2 describes their synthesis into corpus stylistics and its applications, focussing on the studies most relevant to this project. Chapter 3 discusses the concept of style with reference to the work of Henry James, and particularly includes criticism which relates to the question of his 'difficult' late style.

This research project is then described; the building of a corpus (the *Henry James Parsed Corpus* or *HJPC*) based on Henry James's early novel, *Washington Square*, and his last completed novel, *The Golden Bowl*, is detailed in Chapter 4. The analysis of that corpus with the results it yields is outlined in Chapter 5. Chapter 6 describes a different application of the data contained in the *HJPC*; the central chapter of each novel is examined in much closer detail than is possible for the two books as a whole. Finally Chapter 7 summarises the results of this project, both in terms of Henry James's style and the advantages and limitations of this corpus stylistic approach.

# Chapter 2 Corpus Stylistics

---

## 2.1 Introduction

In order to locate my own project within the field, I summarise in this chapter some of the history of and current major trends in corpus stylistics. The term 'corpus stylistics' is defined in section 2.2. The roots of stylistics are traced to pre-computer statistical explorations of literature and early computer-driven studies. The influence of corpus linguistics as it began to be used for stylistic analysis will then be examined. Subsequently corpus stylistics grew in a number of different directions. Corpus linguistic studies occasionally touched on literary texts and concerns, as in the work of Biber (1988, 1995) and others who use his method of multivariate analysis in linguistics (see section 2.6). Corpus stylistics is used widely for authorship attribution, and the seminal work of Leech and Short (1981) on the presentation of speech and thought in literature has been developed further with corpus methods. The dominant area of corpus stylistics concerns the lexis of literary work, which is illustrated with a number of examples in section 2.9. As corpus stylistics has matured, some critics have used it along with other approaches and theories. Corpus stylistics concerned with the syntax of literary texts, which is the subject of this project, remains quite rare, but some examples are discussed in section 2.11. The limitations of these syntactic studies occasion the need for this project.

While Leech and Short's analysis of a Henry James short story will be described in Chapter 3, in the latest edition of *Style in Fiction* (2007) they also comment on the influence of corpus linguistics on stylistics in the last 25 years. They differentiate a wide variety of approaches using corpus techniques from Louw's semantic prosody (see section 2.5) to Hori's analysis of collocation (see section 2.9.2), including Hoover's revisiting of Halliday's paper on *The Inheritors* (see section 2.11) and Stubbs's examination of Conrad's *Heart of Darkness* (see section 2.9.2).

Leech and Short express some appreciation of what they call the "corpus turn' in stylistics" (Leech & Short, 2007, p. 286) and reject worries that computer analysis is taking over stylistics. They applaud the ability to search rapidly and accurately for such elements as vocabulary, collocations and some simple syntactic categories but explain that more abstract levels of analysis require manual annotation of texts. Leech and Short consider that corpus stylistics is only statistically valid where quite large amounts of text are used. They find limited ways to exploit these methods in their case study of a Powys short story. Their computerised searches mostly follow up a critical intuition rather than finding

something unexpected in a text. Hence they conclude that, however much corpus methods may be useful, a human mind is still needed for text interpretation. Their concerns about corpus size and statistical reliability apply to the use of large natural language corpora as comparators. As I will show in Chapter 5, it is possible to use statistics to analyse purpose-made small corpora with accuracy.

## **2.2 Defining corpus stylistics**

Corpus stylistics brings the methods of corpus linguistics to the practice of stylistics. I use the term 'corpus stylistics' specifically to refer to the study of literary texts. Some researchers in this field use 'corpus stylistics' to refer to literary criticism only; Mahlberg characterises it as a methodology which combines different approaches but is fundamentally 'a way of bringing the study of language and literature closer together' (Mahlberg, 2007, p. 219). Wynne (2006) also implicitly assumes that corpus stylistics is a stylistic enquiry into literary language. Others use the term more generally; Semino and Short include news reports and autobiographies in their book, *Corpus Stylistics* (2004). Therefore some writers may specify literary corpus stylistics when necessary, but I shall not do so.

Corpus stylistics is not a theory but rather a methodology which is capable of different applications. In fact, it would be truer to say that corpus stylistics involves a number of different methodologies, linked through the use of computers and quantitative analysis but quite diverse both in their practice, aims and results.

The idea of foregrounding, and specifically the concept of deviation, remains relevant in corpus stylistics despite their early 20<sup>th</sup> century origins. The research project described in this dissertation can be seen as a study of both internal and external deviation – see sections 5.5 and 6.1. In this chapter, after outlining the early development of corpus stylistics and noting some continuing issues, I will describe works which resemble my own research in terms of describing deviation as well as works which use similar corpus methodology.

## **2.3 Pre-computer studies**

Even before computing and computerised corpora, some literary scholars looked at the language of texts in a quantitative way. Indeed Rudman traces 'the quantitative description of texts' (Rudman, 2006, p. 611) from Sanskrit scholars of approximately 300 B.C. through the ancient Greeks and mediaeval commentators on the Hebrew Bible. Rudman (2006) suggests that Victorian scholarship in this area is usually said to begin with Augustus de Morgan in 1851 who studied word-length as a possible indicator of authorial style (in particular, in the Epistles of St.

Paul). He was followed by T. C. Mendenhall (1901) who used word-length to distinguish the works of Bacon and Shakespeare.

Rudman identifies a particular increase in interest lasting from the 1940's to the 1970's. This mid-twentieth century surge is typified by Williams's early, and extraordinarily labour-intensive, work. Interest had shifted from word-length to sentence-length as a tool to establish authorship, and Williams dates 'the real start of the application of modern statistical methods in this type of work' (Williams, 1970, p. 9) to Yule's 1938 study looking at relative frequency in vocabulary. Williams sees the statistical study of language as useful for studying languages, authors and authorship and the development of an author's style, using data such as word-length and sentence-length, comparison of use of parts of speech, lemmas and words of different origin. He discusses how to choose samples of text for study and how to analyse results statistically, and is particularly interested in how the use of language reveals an author's thinking. 'Undoubtedly the best insight into the mind of a writer by statistical methods is obtained by a study of his vocabulary and its usage-frequency distribution' (Williams, 1970, p. 148).

Williams realised that the enormous quantity of labour required for such work could be avoided by the use of computers but distrusted them.

The task of extracting information is at times tedious, and today much of this can be done by expensive electronic machines, but I do not recommend this method for preliminary exploration. It is while sorting the actual evidence that one notices the difficulties, finds ways of avoiding them, and thinks of new lines of investigation. Also I have yet to see a machine that will sort the different parts of speech, or the different meanings of the same letter sequence. Machines often answer questions, but seldom ask them. (Williams, 1970, pp. 15-16)

Williams predicted modern corpus stylistics with great accuracy. To this day there is still an overwhelming focus on lexical studies. I use sentence length as one way of identifying internally-deviant sentences within a chapter of *Washington Square* and *The Golden Bowl*<sup>2</sup>. Like Williams, Semino and Short notice that the detailed examination needed for coding text can be fruitful, and that has been my own experience when creating the *HJPC*.

Our point here is that it was the analysis of the corpus that showed us in each case what we needed to look for in order to test hypotheses we had arrived at from doing the corpus work itself, and where we might be able to find suitable additional data. To put it metaphorically, the corpus has turned out to be more of a springboard than a straightjacket. (Semino & Short, 2004, p. 227)

---

<sup>2</sup> See paragraphs 6.5.1.1 and 6.6.1.1.



While computers can now perform analyses far beyond Williams's imagining, they still struggle with syntactic analysis and homonyms, and, most crucially, the real analysis must be done by a human literary critic.

In his review of 'statistical stylistics' (Bailey, 1969) Richard Bailey also picks out Mendenhall as an important forerunner of modern stylistics, identifying issues which are still relevant such as the choice of sample size, which stylistic features to study and how to formulate, test and verify a hypothesis. Bailey divides mid-20<sup>th</sup> century statistical stylisticians by their interest in vocabulary or syntax, and within these two groups there are different methodologies. An important development was Zipf's work on word frequency in a number of languages, relating it to word-length, and formulating *Zipf's Law* which provides a mathematical model for the frequency of use of words (as well as having wider applications).<sup>3</sup> Further possibilities for study were the development of a writer's vocabulary over time, and whether there were discoverable rules for a writer's lexicon. Other critics looked at word length or the use of words which were native, borrowed or archaic.

These pioneers prefigured the later development of corpus stylistics, focusing mainly on lexis, methodological problems and even cognitive questions, applications such as authorship attribution, as well as interpretation and evaluation. Thus, when corpus linguistics developed, a path had already been made for its application to literary texts, building on the pioneering work of statistical stylistics. The ground had been laid for the emergence of corpus stylistics many years before the first corpora.

## 2.4 Early computer studies

Some possible goals had been identified in earlier years, but with the advent and exponential growth of computer availability and speed, it was possible to try to realise them.

In a paper from 1989, Bailey reviews the use of computers in literary stylistics, while also critiquing the field. While he notes that an enormous amount of tedious work is now done automatically, so that 'a great many things we now attempt in stylistics could not be sensibly undertaken without the help of the computer' (Bailey, 1989, p. 3), he points out the lack of agreement on what to study and what goals to aim at. Bailey explains that most studies have focused on data retrieval, and particularly on lexis, but this is not literary criticism in his opinion, which he sees as centring on close reading and requiring a theory of literary language. He does not dismiss corpus stylistics methodology but sees it as

---

<sup>3</sup> See <http://www.alicebot.org/articles/wallace/zipf.html> for an informal and very clear explanation of Zipf's law.

requiring theoretical development, which could perhaps be achieved by the study of other computer uses, such as in mathematics.

In response to criticisms of this kind, corpus stylistic studies often include arguments in defence of the use of corpus methodology. More recent descriptions of corpus stylistics simply include subjective and qualitative analysis as part of the theoretical description and the practical application of corpus stylistics. For example, Carter is careful to include both objective and subjective approaches in his definition of the field:

[C]orpus stylistic analysis is an essentially quantitative procedure and involves an assessment of significance drawn statistically from a corpus-informed count. The actual application of corpus stylistics to texts necessarily involves ... qualitative decisions and interpretive acts made by the analyst in the light of and to some degree in advance of the results from the assembled data-bank. Corpus stylistic analysis is a relatively objective methodological procedure that at its best is guided by a relatively subjective process of interpretation. (Carter, 2010, pp. 66-7)

This clear summary encapsulates my own approach to corpus stylistics.

At the 1958 conference on style discussed in section 1.2.1, John Carroll, a psychologist with an interest in language, discussed the limitations of literary criticism. 'Literary criticism today does not have any well and sharply defined set of elements by which a sample of prose may readily be characterized' (Carroll, 1960, p. 283). Starting from this viewpoint and at this early date in the development of the field, he was enthusiastic about the possibilities of using computers in literary studies. He suggested that factor analysis, which was used widely in psychology but not in literary studies, would be a useful tool for looking at texts. He had compiled a corpus consisting of 150 passages of various sources and styles of English prose and conducted an analysis covering both objective and evaluative viewpoints. As an objective analysis, he had counted such elements as word classes, clauses and sentences. For the evaluative approach, he had asked eight experts to evaluate his texts on 29 descriptive scales, for example from meaningless to meaningful. He thus arrived at 29 evaluative scores for each text which, added to the objective scores, gave him a total of 68 scores per text. He had then applied factor analysis and identified seven dimensions of literary style (one of which he later discarded as meaningless). 'All these computations were performed with the aid of high-speed electronic computing machines' (Carroll, 1960, p. 288). The content of these dimensions ranged from totally subjective to totally objective. At the end of this analysis, he felt that it was possible to characterise a particular passage but was concerned as to whether he was really measuring style or just types of content. He had tried to get his judges to differentiate between content and style, but found that they could not do so. This early study in corpus stylistics resembles Biber's

work, discussed in section 2.6. It also prefigures the strong growth of cognitive stylistics in recent years.

In contrast, Louis Milic (1967) sees the use of quantitative methods as a way to escape from the subjectivity of the critic, and from the difficulty for the human reader of detecting subtle characteristics of an author's style. This approach requires researchers to use statistical methods to look at texts, comparing them with norms (but leaves unanswered the problem of what norms should be applied in such studies.) Milic illustrates his approach in *A Quantitative Approach to the Style of Jonathan Swift* (Milic, 1967). He encodes features such as parts of speech, gerunds, foreign words and numerals by hand in a method redolent of the early days of computing. 'Selected samples were hand-coded into the two-digit codes and the resulting numerical strings were punched onto punch-cards' (Milic, 1972, p. 266). He attempts to identify particularly Swiftian features but runs into various difficulties. It is a slow, labour-intensive process which imposes a small sample size, and the advance choice of what to code means that only those features can be found. Milic acknowledges that the attempt to avoid subjectivity founders on this problem of choosing coding according to an already-held theory of what might be found and what might be significant. However, he is able to identify features of Swift's style and make claims about Swift's authorship of a disputed work.

In an attempt to escape from this problem, Milic invented a computer program which, though not a complete parsing, 'would analyze a sentence, classify the words in it, identify its constructions and assign syntactical roles and functions to words and constructions' (Milic, 1972, p. 267). At the time of writing his paper, Milic was still working on this program but saw it as a necessary step in the attempt to assess literary texts accurately. It remains a rare attempt to include syntax in corpus stylistics, and, as is explained in section 4.3, the development of a truly effective parsing program remains incomplete.

Over 30 years later, Peter Wilson's paper (2004) on the idiolect of Alfred Jingle, a character in *The Pickwick Papers*, also uses a hand-coded corpus method. Wilson quantifies the extent to which Jingle's speech is marked by 'syntactic incompleteness' (Peter Wilson, 2004, p. 79) and particularly by ellipsis. Angus Wilson (1968) claims that Jingle's unusual idiolect is a 'prefiguration of Joycean linguistic experiments' (Angus Wilson, 1968, p. 34). Peter Wilson tests this hypothesis by analysing a short passage of *Ulysses* with the same classification system as he has used for *The Pickwick Papers*. He finds a number of differences and concludes that

Jingle can be seen to pre-figure Bloom in the syntax of what is expressed, in the formal residual structures of speech and thought representation. What is not expressed, however, the elliptical gaps of Jingle's dialogue and Bloom's interior monologue, are two very different kinds of textual absence. (Wilson, 2004, p. 91)

Without using computer analysis, and though working on relatively small samples of text, Wilson has been able to characterise this example of Dickens's use of speech representation, and to evaluate earlier critical analysis. His quantitative methodology has proved rewarding.

An even more mathematical approach was taken by Wilhelm Fucks, an Austrian physicist who was also involved in early quantitative linguistics. Fucks (1952) developed a formula to work out the frequency of words with different numbers of syllables. Using this he found that *Othello* had 78.81% of 1-syllabled words, 15.11% of 2-syllabled, 4.95% of 3-syllabled, and 1.16% of 4 syllable-words. He compared these frequencies with works by Galsworthy, Huxley and other authors, some of them German, finding that Huxley's writing had the longest words and *Othello* the least. He also calculated the average syllable number per word, claiming that 'a particular work of a particular author is associated with a single characteristic number. Furthermore, this average syllable number per word is peculiar to the author' (Fucks, 1952, p. 125). This seems a doubtful claim and would certainly require a large number of works by each author to be examined to establish it. Also with Shakespeare he was analysing a play in early modern English which would surely in any case be different from the relatively modern prose of Galsworthy and Huxley. Fucks goes on to calculate the distance between words of the same syllable length and comes up with a very stable figure across the different texts.

Fucks's conclusions may be mathematically accurate but illustrate various problems with this kind of methodology. Many variables may affect the style of a text and not all of them are within authorial choice, such as period, genre, etc. Even if this is overcome, there is the vital and ever-present question of the value of the quantitative result. An ability to characterise *Othello* mathematically does not immediately help us to understand or appreciate it; that requires further interpretation by a literary critic.

## 2.5 The influence of corpus linguistics

With the development of corpus linguistics came the possibility of going beyond mere counting to more sophisticated ways of looking at literature quantitatively. Wynne discusses the fusion of the two fields of corpus linguistics and stylistics in his encyclopaedia article (Wynne, 2006). He defines stylistics as 'a field of empirical inquiry, in which the insights and techniques of linguistic theory are used to analyse literary texts' (Wynne, 2006, p. 223). This has resemblances with the similarly empirical field of corpus linguistics which 'relies on the evidence of language usage as collected and analysed in corpora' (ibid. p. 223). Wynne cites Burrows (2002b) as also seeing a resemblance between the two fields. Although the use of corpus

linguistic methods for stylistic criticism is growing, he notes that it is still comparatively rare, although literary corpora of various periods and genres are becoming available as well as large linguistic corpora which may include literary elements. Particular linguistic features can be studied by annotating a corpus in such a way as to specifically highlight those features, which are then available for electronic searching. (Such annotations are called tags.) Pre-existing tagging in a corpus, for example part-of-speech labels, may also be utilised. An example of this approach is the work of Leech and Short (1981), described along with Short's development of the methodology in section 2.8 below. The *HJPC* required annotation, as described by Wynne, but some part-of-speech labels were generated automatically, though they required checking. My methodology is described in detail in Chapter 4.

An alternative approach described by Wynne is to use a reference corpus of natural language as a comparison to a literary text. In this model, literary style is seen as a deviation from the language as a whole, as represented by the reference corpus, that is, it is an external deviation as explained in section 1.2.2. A particular word, phrase or structure can be checked against a wide variety of genres, speakers/writers, etc. which large natural language corpora contain. If a word, phrase or structure is unusually frequent, this might reveal foregrounding. However foregrounding only occurs if the deviation is considered to be significant in revealing literary meaning or effect, adding to a reader's 'appreciation of its contribution to the text as a literary work' (Leech, 2008, p. 163). As Leech points out, this moves the focus from the text to a reader's response to the text, which is only relevant if they are affected, consciously or unconsciously, by the deviance revealed by the comparison with a reference corpus.

Wynne addresses the scarce use of computer techniques in literary analysis, suggesting that literary critics may feel that they lack the necessary technical skills, and that good quality electronic texts are often unavailable and may be difficult to compile because of intellectual property rights. Also, there may be a general resistance to 'scientific' approaches to the arts.

Although it can be argued that the use of computers for analysing electronic versions of texts, and for establishing evidence of linguistic norms in language use, is merely a means of verifying and refining empirical statements and findings, some see the danger of research becoming preoccupied with computational procedures, and the encoding and annotation of electronic texts, leading to a regrettable lack of attention to textuality and the meaning. (Wynne, 2006, p. 225)

Despite such barriers, Wynne is optimistic about the growth of corpus stylistics. Technical difficulties are being tackled, and although there will be those who eschew the use of computers in literary analysis, Wynne asserts that 'corpus

linguistics will prove to be a useful addition to the stylistician's toolkit' (Wynne, 2006, p. 226).

Louw (1993, 1997) has taken advantage of the availability of large natural language corpora in his work, which develops Sinclair's notion of semantic prosody. (The germ of the idea of semantic prosody can be seen in Sinclair's description of a 'lexical item' (1966):

a formal item (at least one morpheme long) whose pattern of occurrence can be described in terms of a uniquely ordered series of other lexical items occurring in its environment. (Sinclair, 1966, p. 412)

In the same paper Sinclair also discusses the notions of a node with a span, that is 'the number of lexical items on each side of a node that we consider relevant to that node' (Sinclair, 1966, p. 415).

Semantic prosody is defined by Jeffries and McIntyre as 'the connotations that [a word] takes on as a result of the meanings of the words that it collocates with' (Jeffries & McIntyre, 2010, p. 184). The essence of the concept is that a word takes its meaning, at least partly, from the sequence of words in which it occurs; such sequences can be found in a corpus with the use of concordancing software. Louw quotes Sinclair at the Leeds Conference on Phraseology in 1994:

One hypothesis is that the notion of a linguistic item could be extended, at least for English, so that units of meaning were expected to be phrasal, and the idea of a word carrying meaning on its own was relegated to the margins of linguistic interest, in the enumeration of flora and fauna for example. (John Sinclair quoted in Louw, 1997, p. 247)

Semantic prosody, Louw feels, is an extremely important concept for both literary criticism and pedagogy. In his much-cited 1993 paper (Louw, 1993), Louw explains that semantic prosody 'relies for its effect on a collocative clash which is perceived, albeit subliminally, by the reader' (Louw, 1993, p. 157). This 'collocative clash' comes about when there is an expected collocation which is violated, although Louw explains that these collocational expectations are not available through introspection. They can, however, be found using computer analysis. The idea is illustrated in an examination of Larkin's poem *First Sight* (Louw, 1993). Using COBUILD,<sup>4</sup> Louw looks at collocates of the word 'utterly' which appears in the last line of the poem, finding that it has negative connotations; this gives the line a note of foreboding, although the poem does not state this explicitly. This, Louw claims, is how poetry conveys meaning to the reader using concentrated language which has the capacity to give a particular resonance to a line succinctly.

---

<sup>4</sup> The Bank of English, also sometimes called COBUILD, is currently a 450 million word corpus of present-day English with a subcorpus for teaching purposes consisting of 56 million words. See Introduction section of <http://www.titania.bham.ac.uk/docs/svenguide.html> (accessed 12.12.2013)

Louw explains that authors also develop their own collocates, making their own system of symbolism within their writing. This is a disruption of the collocational norm familiar to the reader, a particular kind of external deviation. An example is Yeats, who in his poetry collocates 'ladder' with 'stair' and 'dance' when the most common collocates are 'rungs' 'climb' and 'social'. 'This phenomenon appears not to have been examined by literary critics and the computational evidence for it now needs to be laid before them' (Louw, 1997, p. 248). Louw concludes with a strong statement of the importance of corpus stylistics.

If it can be demonstrated that certain forms of literary knowledge might never have been apprehended were it not for the use of computers and corpora, then it follows, probably to an extent yet to be determined, that our critical theory ought to provide a permanent role for corpora in criticism. (Louw, 1997, p. 248)

Though critiqued by Hunston (2007) Stewart (2010) and Philip (2011) among others, semantic prosody continues to be an important concept within corpus stylistics, with a steady stream of papers printed on the subject. Such studies are facilitated by corpus search tools which readily provide concordance data.

## **2.6 Multidimensional Analysis**

With the development of corpus linguistics, research in that field sometimes includes or spills over into stylistics. An example is Biber's (1988) development of his method of multidimensional analysis, which is used to identify the linguistic characteristics of different spoken and written language registers. The method is explained in a co-authored book a decade later (Biber, et al., 1998). Biber et al. identify the difficulty of knowing which linguistic features are significant in distinguishing different registers, as there is a vast range of possibilities. The initial choice of features to analyse is made by reviewing existing research, including the widest and most inclusive possible range. Once the features have been identified, the texts to be studied are tagged automatically, finding all the target features and calculating the frequency with which they occur. 481 texts were analysed in this way, comprising 960,000 words from the LOB corpus and the London-Lund corpus, both of which include information on the register of their texts. The core of Biber's method is the identification of co-occurrence of linguistic features.

[W]e need to understand the ways in which linguistic features work together - or "co-occur" - in texts. From a linguistic perspective, registers are distinguished by the extent to which they utilize such different sets of co-occurring features. (Biber, et al., 1998, p. 144)

The co-occurrence of features is further analysed using factor analysis 'to identify sets of variables that are distributed in similar ways' (Biber, et al., 1998, p. 146). This procedure is 'based on the assumption that frequently co-occurring linguistic

features have at least one shared communicative function' (Biber, 1988, p. 63). These groups of co-occurring features are then labelled as a "'dimension' of variation" (Biber, et al., 1998, p. 146) which is given a functional interpretation 'through assessment of the communicative function(s) most widely shared by the features constituting each factor' (Biber, 1988, p. 64 in Table 4.1). The dimensions have two poles and texts can be assessed to see where they fit on that dimensional axis. For example, Dimension 1 separates texts on an axis of 'involved versus informational production' (Biber, et al., 1998, p. 151). Informational texts, such as academic journal articles, are those with longer words, more nouns, and more nouns modified by attributive adjectives or prepositional phrases as well as other measures. 'Involved' texts are typically face-to-face interactions and are marked by a low occurrence of the same syntactic features.

Biber et al. apply this method of analysis to many types of texts, both spoken and written. They do not focus on literary criticism but some literary texts are used. When Biber et al attempt to characterise an author's style, they identify two problems: it is difficult to identify a comparative framework for the text and to ascertain which linguistic characteristics should be studied (echoing Milic (1972)). Two passages from Johnson's *The History of Rasselas: Prince of Abyssinia*, one narrative and one descriptive, are compared to 33 fiction texts over four centuries, all from well-known authors. The results show Johnson's descriptive passage standing out clearly from all the others in three dimensions, with characteristics more like informational prose than fiction. The text is 'extremely informational', 'extremely elaborated in reference' and 'extremely impersonal in style' (Biber, et al., 1998, p. 224). It appears that Johnson's style is very idiosyncratic in comparison to his contemporaries and to writers of other periods.

Later critics have applied multidimensional analysis for literary analysis. Examples are Geisler (2002), Andreev (2008) and Craig (1999). Although Biber et al's methodology is very different from my own, it resembles my project in the use of syntactic features to characterise a literary text. Multivariate analysis is a productive and powerful approach but is not designed as a mode of literary criticism. Though it can be adapted for that purpose, as Craig has done, the focus is on functionality and so the critical focus is shaped by the method.

## 2.7 Authorship attribution

Corpus stylistics has been particularly widely used to tackle problems of authorship, both literary and otherwise. In this case, it is often called stylometry or stylometrics. However, this use of corpus stylistics, and particularly the exact way methodologies are applied, can be controversial. Rudman points out that, when used for authorship identification, corpus methodology is based on the hypothesis



that 'every writer has a unique and verifiable style' (Rudman, 2006, p. 611). He lays out guidelines for a well-conducted authorship study, including matching texts for genre, period, etc. and careful preparation of the texts. He notes that there are hundreds of thousands of quantifiable stylemarkers at every level from the paragraph to below the word. It is difficult to know which stylemarkers to count and then there is a bewildering choice of statistical methods to apply, some of which are criticised by other statisticians. Rudman also suggests that it is hard for literary academics to become sufficiently expert at statistics to use such programs reliably. (It may be better for statisticians to turn to literary analysis.)<sup>5</sup>

In a paper utilising multiple methods for authorship attribution, Burrows' careful methodology, which he explains and justifies at each point, answers some of Stevenson's (1989) criticisms. His work is based on his (later formulated) belief that

[e]vidence of authorship pervades whatever anybody writes. Provided appropriate procedures are employed in the analysis of an appropriate set of texts, it can almost always be elicited. (Burrows, 2007, p. 28)

He examines the issue of whether the "Painter Satires" discovered in 1945 are by Andrew Marvell. This question has been examined by other experts on Marvell using their knowledge of the period and comparing the Satires with the work of Marvell and others stylistically. Burrows offers computational stylistics as 'an entirely fresh approach' (Burrows, 2005, p. 282). He sets out to test whether 'any of these poems exhibit a closer stylistic affinity for each other than for any of the others' (ibid. p. 282) by comparing them to 25 other long poems of the late 17<sup>th</sup> and early 18<sup>th</sup> centuries. Using standardised word-counts of the 150 most common words, as calculated from a database of 17<sup>th</sup> century poetry, a word-frequency profile is obtained for each text. These profiles are then subjected to cluster analysis, which Burrows explains is particularly useful when the question is open as to whether the texts resemble each other. It also produces very easy-to-read 'family trees' of the texts analysed.<sup>6</sup>

The figure illustrating the result shows that the *Second Advice* and the *Last Instructions* are most closely related, and the *Third Advice* resembles them. The *Fourth Advice* and the *Fifth Advice* are far more separate. Burrows has included a text from a different genre in his study and can therefore illustrate that 'genre has less influence than authorship upon this outcome' (Burrows, 2005, p. 287). From this it would seem that the *Second* and *Third Advices* and *The Last Instructions*

---

<sup>5</sup> C.f. Stevenson (1989), who also critiques computational stylistics, particularly authorship studies, highlighting the difficulties of using statistical tests correctly and the danger of drawing unwarranted conclusions from the data.

<sup>6</sup> Although he does not use this term, Burroughs has, in effect, calculated the amount of deviance between his texts.

have the same author, and that the *Fourth* and *Fifth* were written by other people. Further analysis results in the three *Advices* which formed a cluster showing similar characteristics to the known Marvell poems.

To confirm the analysis Burrows tries to retest in a different way using a larger database of 25 poets and his own "Delta" procedure.<sup>7</sup> Delta involves compiling a frequency hierarchy of the most common words in a large group of texts. When applying the Delta method to the 'Painter Satires', Burrows looked at each poem in turn asking the question 'From which of our twenty-five authorial profiles does the corresponding profile from this poem (or this) diverge least?' (Burrows, 2005, p. 287) Marvell has the closest profile to the three *Advices* which have clustered together previously; others have a far lower resemblance.

Burrows then looks at style more directly. In comparison to the other 25 authors, Marvell uses some words much more frequently and others much less. Burrows picks out his frequent use of 'the', his low use of post-modification of nouns and common prepositions, stating that 'the overall effect is of an unusually strong emphasis upon named things' (Burrows, 2005, p. 254). Added to his low use of negatives and auxiliary verbs, this makes for 'an energetic, closely specified, and subtly discriminating poetic style [which] may suggest how deeply such characteristics are embedded in the common stuff of the language' (Burrows, 2005, p. 295).

Using principal component analysis,<sup>8</sup> Burrows reconfirms his previous finding, showing in scatterplots that the three *Advices*, which now appear to be confirmed as Marvell's work, fit exactly with his own work in these strong preferences and avoidances of some common words.

Burrows's work is unusual in his interest in part of speech categories, focusing more on the syntax of common words than their meaning (and hence is closer to my methods than most stylometrics research). His mastery of statistical methodology allows him to combine tests which produce robust results, demonstrating not only a way of addressing particular authorship questions but also the usefulness of the corpus stylistic approach.

Burrows revisits the 'Painter Satires' to test his more recent Zeta and Iota developments of Delta (Burrows, 2007). The new tests utilise low frequency words rather than the high frequency lists of Delta. Each uses word lists from texts from a possible author of a text (the 'base-set') and a multi-author reference set (the 'counter-set'), and contrasts are set up between the two lists. Burrows has some

---

<sup>7</sup> Burrows's definition of a delta-score is 'the mean of the absolute differences between the z-scores for a set of word-variables in a given text group and the z-scores for the same set of word-variables in a target text.' (Burrows, 2002a, p. 271)

<sup>8</sup> Craig describes principal component analysis as 'a method for "data reduction", aiming to find a handful of factors which account for most of the variation in a given set of data.' (Craig, 1999, p. 222)

success with distinguishing works from the 'Painter Satires' written by Marvell, but where there is more than one possible alternative author his method is less successful.

Burrows's methodology is applied to literary texts by David Hoover, exploring further the possibilities for author attribution as well as gaining some insight into poetic style. Hoover applies Burrows's analytic tool Zeta and the later Iota in a study of some modern American poetry. He explains that Zeta and Iota

are specifically designed to focus attention on words that are below the stratum of the most frequent words and that are characteristic of a text or an author. (Hoover, 2008, p. 213)

Hoover finds Burrows's tools very successful at recognising poems and their authors in his corpus of modern American poetry by 25 poets born between 1869 and 1913. In particular, he finds contrasts between the vocabulary used by Stevens and Frost: Frost uses many contractions and Anglo-Saxon origin words 'with a rural flavor' (Hoover, 2008, p. 216), whereas Stevens's vocabulary is more Latinate, more formal and abstract. Hoover notes that Stevens has a reputation for being a difficult poet to read. This exercise leads Hoover to give a more general statement about stylistics which echoes the idea of foregrounding:

This reminds us that studying style is always a comparative undertaking: no feature can be striking or characteristic unless it differs from some norm or imagined alternative. (Hoover, 2008, p. 217)

While the data on Stevens and Frost emerged from a large-scale study, Hoover says that a simpler method can be used with a sample of only two poets; he uses the work of Stevens and Robinson. He makes a word list from a sample of each poet's work and selects any words which are at least three times as frequent in one poet compared to the other. This, he says,

gives a very good contrastive picture, even without the additional information about the consistency of use that is present in Zeta and (to some extent) in Iota. (Hoover, 2008, p. 219)

With this data, Hoover notes that Stevens's edited word list contains almost all content words, often relating to the countryside but including abstract ideas. In contrast, Robinson uses many more function words. The content words he uses are 'much less specific and less striking' and include "'moral' words like envy, praise, pride and worth" (Hoover, 2008, p. 222). He notes that 'it would be fair to characterize Robinson's style as more grammatical and Stevens's as more lexical' (Hoover, 2008, p. 222).

Following up on the success of this simpler process of analysis, Hoover attempts to replicate it using the samples from the 25 poets he analysed with Burrows's tools. He removes the poetry of Stevens from the corpus and then makes a word frequency list for the other samples. He then finds the maximum frequency

for each word in the whole corpus (apart from the poetry of Stevens) and compares that frequency with the same word in Stevens's work. A new list is made of any words which are at least 1.5 times as frequent in Stevens as in any other poet, or which are used by Stevens at least twice and not by any other poet. Hoover then repeats this whole process for Robinson, Eliot and Sandburg. Finally these very characteristic words are reproduced in a table which lacks the poets' names, so that the reader can see if they recognise the lexis of the poets. Hoover debates the usefulness of this process. It certainly reveals which poets use very unusual words and which, like Robinson, use common words. However, Hoover himself is unclear whether mere recognition is a worthwhile area for investigation. He concludes by suggesting that this procedure reveals words which could then be used for further exploration of the style of one or more poets.

Hoover's work also calls into question the necessity for the complexity of Zeta and Iota. It may be worthwhile for establishing authorship but for stylistic analysis Hoover's simplified procedure was successful in isolating characteristic lexical choices. However, that should only be a first step in a more qualitative and evaluative criticism.

## **2.8 Speech, thought and writing**

One of the strong strands which have emerged in corpus stylistics is research into the representation of speech, thought and writing. A key influence has been Leech and Short's seminal book, *Style in Fiction* (1981). Short develops his framework for the analysis of speech, thought and writing in his collaborative work with Semino (Semino & Short, 2004). Here the authors experiment with corpus stylistic methods to re-examine their previous work in discourse presentation. They say that their early work, like all theoretical models in stylistics up to that point, was developed through the use of scholarly intuition, based on extensive personal reading experience, which was in turn exemplified and tested through the analysis of examples chosen from previous reading (Semino & Short, 2004, p. 4). The authors emphasise that the corpus work they do is only one of the tools to be used, but that it is useful in a number of ways. In contrast to some critics (they cite Fludernik, 1993), they found that the actual work of annotating the corpus was a useful spur to analysis.

For this study, Semino and Short use the framework of speech and thought presentation presented in Leech and Short (1981). They compile a corpus of '120 text samples of approximately 2,000 words each' (Semino & Short, 2004, p. 19), of which approximately one third was prose fiction. Prose fiction was then sub-divided first into 'serious' and 'popular' works, and further into first- and third-person narration. The corpus had to be tagged by hand and included labelling with a

spectrum of possibilities from Free Direct Speech to Narrator's Report of Speech and the equivalents for thought and writing. In applying this tagging system, Semino and Short found that various difficulties were highlighted, which led to some modification. As the corpus was tagged by hand, human error was unavoidable, although multiple checks were made. In addition, they acknowledge that some decisions are arguable and could still be debated. By the end of the study, Semino and Short were able to list some characteristics of each of their genres and were able to summarise quantitative differences between fiction, newspaper writing and autobiographical writing.

With some caveats, Semino and Short end on a note of approbation for corpus stylistics.

The corpus stylistics approach has not prevented us from doing anything we would have done before ... but it has enabled us to find out a great deal more than we would otherwise have been able to do. That said, the kind of corpus-based work we have done involves a considerable amount of time and resources, and can only be realistically undertaken if financial support is available. (Semino & Short, 2004, p. 226)

## 2.9 Lexical studies

The great majority of corpus stylistic studies focus on the lexis of their chosen texts. Carter (2010) argues that this dominance is inherent in the procedures of corpus linguistics and in particular the development and use of very large corpora.

In a basic sense, corpus linguistic description of language prioritizes lexis. Whereas stylistics pays more attention to deviations from linguistic norms that lead to the creation of artistic effects, corpus linguistics focuses on what can be identified computationally – which tends to be on lexical patterns, especially patterns that are frequently repeated. (Carter, 2010, p. 65)

Within this general approach, some researchers focus on particular kinds of lexical search.

In the first chapter of her edited collection, Archer (Archer, 2009) explains the basics of lexical corpus stylistics. At its core, the methodology involves making a word list which is then analysed. There are many ways in which the list can be analysed and many reasons for doing so. Author attribution may be in question, as discussed above. For corpus stylisticians a key point is that patterns can be discerned which are not discoverable by a human reader. The question of the validity of using patterns which are not discernible by the reader runs through much of corpus stylistics. It may be that the reader notices such patterns subconsciously. If not, their role in the text and the validity of their use as part of a critical analysis needs further explanation and justification.

Once a word list has been compiled, further significant patterns can be identified using a concordance, which can show collocates and/or colligates. The

word list can be compared to that of a large natural language corpus as a way of discovering unusual frequencies of words in the text. As function words are extremely frequent, they are usually excluded but some studies, like Burrows's work described in section 2.7, may make use of function words specifically. However, ultimately there must be interpretation by the literary critic.

### 2.9.1 Frequency

A simple way to look at lexis is to compute the frequency with which certain words appear. An example is Burrows's (1986) study of Jane Austen's characterisation, which precedes his work on Delta, Zeta and Iota described in section 2.7. Here he emphasises the many new possibilities that computing had brought to stylistics at that time. By not only picking out, but also analysing, the lexis of novels according to word-type, frequency, speaker, etc., Burrows suggests that the idiolects of fictional characters can be compared. He focuses on Austen's characters and says that Elizabeth Bennet and Emma Woodhouse are most alike and that it is possible to pick out the characteristics of 'authoritarians' or 'vulgarians'. Burrows focuses particularly on modal verbs and on personal, possessive and reflexive pronouns. As so often with corpus stylistic studies, he particularly stresses the importance of tying statistical results to meaning and the proper concerns of literary criticism, such as characterisation. For example, he theorises that the use of 'must' and 'ought' betrays a moral attitude in a character.<sup>9</sup>

Culpeper (2002) analyses Shakespeare's style in *Romeo and Juliet*, focusing on lexis and using corpus stylistic methods. He is interested in the portrayal of character which he hopes to locate in the use of 'style-markers' defined as 'words whose frequencies differ significantly from their frequencies in a norm' (Culpeper, 2002, p. 13). Culpeper explains that stylemarkers are the same as key words as revealed by WordSmith Tools (Scott, 2004). He uses them to compare the main characters in *Romeo and Juliet* (Romeo, Juliet, the Nurse, Friar Lawrence, Capulet, Mercutio, and Benvolio). He does not use an external, natural language corpus but makes a corpus of his own by annotating his electronic text of *Romeo and Juliet* (in modern spelling) with code to indicate dialogue and the speaker. (However, he does include in his comparison table a list of the top 10 word frequencies in both present-day written and spoken English.)

Culpeper's methodology allows him to make statements about how Shakespeare has differentiated the chosen main characters from the other dramatis personae. Using the keyword search and, like Burrows, including function words, he notices that 'Mercutio's top four words are identical to the words for present-day

---

<sup>9</sup> Burrows explores both Jane Austen's prose and this methodology further in *Computation into Criticism* (Burrows, 1987)

written English and Friar Lawrence's top four words appear in the top six for present-day written English' (Culpeper, 2002, p. 17). From this he suggests that

Romeo, Juliet, Capulet and the Nurse are more interactive characters than Mercutio and Friar Lawrence, who both tend to extol forth regardless of other characters on stage. (Culpeper, 2002, p. 17)

Culpeper comments that Mercutio has an elaborate style which might be like writing but it remains unclear why these two characters should be closest to modern-day English.

The keyword lists generated for each of the six main characters are found to reflect various aspects of their concerns or character. For example, Romeo uses words which fit with the lover's role and his references to his own body are considered to be a sign of egocentricity. Culpeper explains that while some of his data gives expected results, other data offers unexpected but understandable results, such as Juliet's use of 'if' and 'yet', reflecting her uncertain state through much of the action of the play. He also points out that these are non-content words which many corpus stylisticians remove from their data but which he has found revealing. Some keywords need examination to reveal their full import. For instance, Capulet has 'go' as a keyword. On examination, this is found to be an imperative and, with other imperatives, reveals his commanding role and character.

Although this study is on a different period and genre to my own, it resembles my work in its focus on the use of corpus methods to reveal the use of syntax in the making of literary style. It also illustrates the goals of corpus stylistics: to find some data which could not be found easily, or perhaps at all, by reading and to use such data to explore the style and/or effect of a literary text. Culpeper subsequently extended this project to include the generation of key part-of-speech and semantic category searches (Culpeper, 2009). He found that, while a keyword search was the most productive procedure, the production of the key semantic categories used by each of the six main characters in *Romeo and Juliet* also produced some unexpected results. For example, Romeo's speech was found to include many words in the category 'Colour and Colour Patterns', showing his use of such words both literally and metaphorically. Culpeper concludes that this reveals the complexity of Romeo's character (in contrast, for example, to the Nurse), as well as, possibly, his higher status. However, Culpeper cautions that semantic tagging in Wmatrix requires careful checking by hand, as the reliability of the tagging is relatively low.

Archer, Culpeper and Rayson's (2009) paper on Shakespeare's tragedies and comedies also adds the extra dimension of semantic analysis to the familiar keyword approach. They identify keywords within key semantic domains which have been automatically chosen, and state that this allows them to find keywords which would not have come to the fore in a standard keyword search. This, they

contend, gives an enhanced view of the content of the text. Their main focus is on love within three comedies and three tragedies, finding differences between the two sets. The tragedies have relatively less occurrences of *love* in the context of 'intimate sexual relationships' and 'liking' (Archer, et al., 2009, p. 156). The 'love-tragedies' focus on other domains and when love is represented "it is much 'darker', and may typify 'tragical' love as opposed to 'ideal' or 'romantic' love" (Archer, et al., 2009, p. 156).

Archer et al.'s methodology bears some resemblance to my own, insofar as it compares data within an author's works rather than comparing data to a natural language corpus or other reference corpus, although their focus is purely on lexis.

### 2.9.2 Collocation

A popular way of using corpus stylistics to examine the lexis of a text is to focus on collocation. Stubbs summarises the rationale for this practice.

Words have a tendency to co-occur with certain other words, and culturally and communicatively competent native speakers of English are aware of such probabilities and of the cultural frames which they trigger. (Stubbs, 2001, p. 17)

In a later paper (Stubbs, 2005), he explores the ramifications of this idea, which has a close resemblance to semantic prosody<sup>10</sup>, looking at the frequency and distribution of individual words and lexemes to find underlying themes in Conrad's *Heart of Darkness*. Stubbs sums up his justification for this pithily, saying that 'collocations create connotations' (Stubbs, 2005, p. 14). He aligns the goal of his project with Kenny (1992) who, he says, sets out the need for 'results which would be impossible to obtain without a computer, and ... [can] be respected as an original scholarly contribution within literary studies' (Stubbs, 2005, p. 5). However Stubbs also states that 'even if quantification only confirms what we already know, this is no bad thing' (Stubbs, 2005, p. 6) as this correlation between quantitative and qualitative analysis could give confidence in the new methodology.

Applying his approach to explore the theme of vagueness in *Heart of Darkness*, Stubbs compares the frequency of words such as *somebody*, *sometimes*, etc. with two reference corpora. One corpus, which he denotes FICTION, comprises the 'imaginative prose' sections of the Brown, LOB, Frown and FLOB corpora<sup>11</sup> and

---

<sup>10</sup> The difference between collocation and semantic prosody is debatable, and may not exist at all. I suggest it is simply a question of emphasis. The notion of semantic prosody emphasises pragmatic meaning when words are found to collocate. A discussion of collocation may emphasise the juxtaposition of words first before looking at the created meaning.

<sup>11</sup> 'Brown, LOB, Frown and FLOB are parallel reference corpora of one million words each of written English. The Brown corpus (prepared at Brown University under the direction of W. Nelson Francis and Henry Kučera) comprises written American English, published in 1961. LOB (prepared at Lancaster, Oslo and Bergen Universities, under the direction of Geoffrey Leech and Stig Johansson) comprises written British English, published in 1961. Frown and FLOB (prepared at Freiburg University under the direction of Christian Mair) are comparable corpora, of written American and British English, published in 1991. The BNC sampler consists of one million words each of spoken and written British English, extracted from the 100-million-word British National Corpus.' (Stubbs, 2005, p. 22 Note 2)



contains over 710,000 words. The other, WRITTEN, is the written section of the BNC, and includes over one million words. These comparisons show that 'frequencies [of words connoting vagueness] are consistently higher in HEART [i.e. *Heart of Darkness*] than in FICTION, and higher in FICTION than in WRITTEN' (Stubbs, 2005, p. 10). Stubbs acknowledges that frequency lists may not accord with what readers find significant, or even notice, but considers them a good starting point, with verbs offering more stylistic information than other parts of speech.

Stubbs goes on to explain that it is useful to go beyond simply looking at lexical frequency. The search for collocations can be used to illuminate the juxtaposition of thematic words. For example, *grass* is often associated with death and decay, and Stubbs produces a collocation table from the novel illustrating this. This raises again the criticism that the researcher decides from his own reading, or the work of other critics, which words to search for in collocations. Stubbs counters this by suggesting that it is possible to 'identify clusters of words which co-collocate across the text, by recording the collocates of each word in the text within a given span, and then summing the collocates for each node-word' (Stubbs, 2005, p. 15). This method is not entirely clear from Stubbs's paper but presumably the software produces a collocation list centred on each word of the novel, which is then the node-word of the collocation (he excluded the 10 most frequent words, which are all function words). The collocations for some node-words are then brought together, based on their connected meanings. Stubbs identifies what he considers to be a genuinely objective set of collocates. He finds a set of words including *still/stillness*, *gloom* and *brooding*, but it seems unlikely that these were the only set which the process threw up, so that the researcher still has to pick out the sets he wishes to discuss from the computer-generated collection.

Stubbs also stresses the importance of Conrad's use of 'recurrent lexicogrammatical patterns' (Stubbs, 2005, p. 15), which may have given rise to the criticism that *Heart of Darkness* is repetitive, when it is possible to show by comparison with other novels of the period that his vocabulary is not particularly restricted. These patterns are often found to have negative meaning in *Heart of Darkness* to a greater extent than in the reference corpora. An example for illustration is 'the darkness of an impenetrable night' i.e. two nouns in a noun phrase with a modifying adjective with a negative meaning. Stubbs goes on to explore other sequences which are found by frequency searches. However, the choice of what to look for, or which sequences to select for analysis and discussion, is consistently informed by Stubbs' own interpretation of the novel or the work of other critics. Stubbs recognises this:

---

I am claiming not that quantitative corpus methods produce entirely new insights into the text, but that they describe more accurately the range of lexico-grammatical patterns which Conrad uses. (Stubbs, 2005, p. 19)

Stubbs finds that comparison of Conrad's language with a carefully-chosen relevant large reference corpus enables him to identify external deviation in *Heart of Darkness*, where Conrad's language is foregrounded by its unusual form. He also finds the reverse, that is phraseology which is familiar and frequent in other texts. He speculates that this may contribute to the enduring popularity of the novel. Stubbs rejects the idea that this undermines his methods, stating that computer analyses 'sometimes document more systematically what literary critics already know (and therefore add to methods of close reading), but they can also reveal otherwise invisible features of long texts' (Stubbs, 2005, p. 22). He concludes:

the aim is to say systematically and explicitly what something is: and that is where empirical, observational analysis can contribute. It is not possible (or desirable) to avoid subjectivity, but observational data can provide more systematic evidence for unavoidable subjective interpretation. (Stubbs, 2005, pp. 21-22)

Mahlberg has written extensively both on the theory and practice of corpus stylistics with a special focus on collocation, particularly in the works of Dickens. While her work is in some ways similar to that of Hori (2004) discussed below, she prefers to include function words, which he excludes. Like many others, she uses Mike Scott's *Keywords* program from his software suite *Wordsmith Tools* (Scott, 2004). Mahlberg says that 'the artistic effect of a text is something that is noticeable' (Mahlberg, 2007, p. 223), an idea which resembles the concept of foregrounding, and is relatively reader-focused rather than text-focused. However, a reader may notice an effect without being able to detect exactly its source. Mahlberg claims that corpus stylistics can elucidate such instinctive critical responses. Her particular focus is on 'local textual functions' which 'characterize (a group of) lexical items with regard to the functions they fulfil in (a group of) texts' (Mahlberg, 2007, p. 224). Her corpus consists of 23 texts, comprising about 4.5 million words which are compared to a similar sized reference corpus of 19<sup>th</sup> century writing, although not all these texts are novels. Mahlberg aims 'to identify local textual functions as stylistic features of texts in the Dickens corpus' (Mahlberg, 2007, p. 224) while agreeing with Leech and Short that the selection of what to focus on in a text is inevitably a subjective critical judgement, and will vary from text to text.

While noting that 3- to 5-word clusters are the usually agreed size, in this study Mahlberg finds 5-word clusters to be useful. The latter are strongly linked to individual texts and often to particular characters. Dickens's use of special turns of phrase to identify and characterise his characters, as well as for humorous or dramatic effect, is well-known and commented on by critics. Mahlberg

acknowledges this and agrees with Stubbs that confirming critical consensus is not a problem for corpus stylistics. On the contrary, she explains that 'a strength of corpus stylistics can lie in the potential of corpus linguistics and literary stylistics complementing each other' (Mahlberg, 2007, p. 228), with the caveat that the corpus stylistician should have an open mind as to whether their research will support or confound previous critical comment.

Mahlberg divides the 5-word clusters into five functional groups, such as those which include a body part noun, or have a naming function, such as *the young man of the name of Guppy*. These are applied in detail to *Bleak House*. She finds that some clusters link characters, illustrating the relationship between people or their parts in the plot. For example, there are similarities in clusters identified with Tulkinghorn and Guppy, who are both trying to find Lady Dedlock's secret. Clusters identified with Esther can be seen to be characteristic of her personality. However Mahlberg warns

[i]t is important to note that the clusters are only pointers to more detailed questions of analysis; clusters alone provide an incomplete picture. (Mahlberg, 2007, p. 236)

Also, the classification of the clusters is pragmatic rather than entirely principled. Mahlberg concludes

[i]t is important to stress that such an approach cannot claim to provide a comprehensive picture of a novel. What it can do is highlight features that are made visible with the help of corpus linguistic tools and that can then form the basis for more detailed critical discussion. (Mahlberg, 2007, p. 239)

In Mahlberg's most recent book, a new concept is added which she calls 'suspensions'. She attributes this to Lambert who defines the 'suspended quotation' as a 'protracted interruption by the narrator of the character's speech' (Lambert, 1981, p. 6). The minimum required length of the interruption is set at 5 words. For example,

"If I understand," *said the Doctor, in a subdued tone*, "some mental shock?" (Mahlberg, 2013, pp. 170, Example (12))[Mahlberg's italics]

Mahlberg employs this idea to explore an adaptation of her Dickens corpus, namely being able to search for suspensions by using speech marks as a guide. A special search program is employed for this purpose, which uses XML annotations added for the purpose.<sup>12</sup> She finds that the suspension is often used to describe the body language which Dickens uses to individualise his characters. These characteristic phrases are given the general name of Labels, whether used in a suspension or not; Mahlberg explains that the term 'Label' relates 'to the more striking repeated

---

<sup>12</sup> The program is called CLiC (Corpus Linguistics in Cheshire) and is an extension of Cheshire 3, which is 'an open source search retrieval engine for XML data' (Mahlberg, 2013, p. 172).

phrases associated with characters that are typically discussed in literary criticism' (Mahlberg, 2013, p. 152). Thus, the concept of suspensions helps to locate Dickens's use of Labels, and hence his method of individualising his characters. Mahlberg's use of the concept of suspensions has resemblances to my own concept of *delay*, which is described in detail in section 4.4.3 and employed in Chapter 5 and Chapter 6. The body of Mahlberg's work is very successful in bringing together corpus methodologies into a genuinely literary corpus stylistics, elucidating Dickens's novels and his literary methods.

Hardy's goal in his paper on O'Connor (Hardy, 2004) is to carry out a literary investigation which would be difficult, if not impossible, without the aid of corpus methods. If this is achieved, it would contrast with Mahlberg's and Stubbs's papers, where each author points out that many of their findings have already been described by conventional stylisticians. Comparing O'Connor's fiction to the Brown corpus, Hardy establishes that the word *eyes* is significantly more common in O'Connor. He takes this to be a sign of O'Connor's theory of sacramentalism, in which everything, animate and inanimate, is seen as imbued with divinity.

Hardy wants to extend this mere confirmation of critical intuition by looking at the contextual use of *eyes*. The most frequent collocates of *eyes* are *his*, *her* and *were* but he dismisses these as uninteresting and then excludes grammatical words. Having derived a collocation table, he examines the grammatical construction of the phrases which include the word *eyes*; he finds that the most common pattern is *eyes* followed by a verb. He suggests that a high frequency of *eyes* + verb 'is a result of, or at least can be interpreted as of a piece with, O'Connor's literary celebration of the sacramentalism of the human body' (Hardy, 2004, p. 418). It is unclear why the combination of *eyes* with a verb should accomplish this more than the second most common collocation, which is *eyes* with a noun-modifier, or indeed that *eyes* + verb has that connotation at all.

Focusing on collocational methodology, Hardy emphasises that it must be supplemented with grammatical analysis in a two-step process, and further that the researcher must consider if the grammatical patterns he finds have semantic significance. He concludes that 'collocational analysis, then, provides us a powerful tool for discovery of stylistic patterns above the level of the individual word' (Hardy, 2004, p. 424). This novel integration of syntax with lexis gives a greater depth to the analysis of collocation in O'Connor's work in comparison with the reference corpus. However, some of the links O'Connor makes between syntax and the underlying and metaphorical meanings of the text are far from self-evident. While the collocational data has been identified objectively, Hardy's interpretation is very much a personal one, and is also conditioned by the previous interpretation of literary critics.

Hori's work (2004) has some resemblance to Mahlberg's discussed above. However, Hori does not use long clusters like Mahlberg but instead uses more conventional short collocational groups. Hori starts with a search for the highest frequency content words in the Dickens corpus, which are then used to find their collocates, as well as considering colligates and semantic prosody. He concludes that 'Dickens has a unique collocational style, which is best revealed through the analysis of usual collocations of common words' (Hori, 2004, p. 56). This is seen in his repeated use of collocates which are unique or little used by his contemporaries, as well as his avoidance of collocates they prefer. Dickens is particularly prone to inventing collocations. In Hori's case study on *Bleak House* he finds more than 700 collocations which are unique to Dickens as well as neologisms.

Hori makes a case for the detailed information about authorial style which this kind of analysis allows. For instance he establishes that the word *love* is unusually rare in Dickens in comparison to the reference corpus but the collocation *my love* is unusually common. These usages, he suggests, are 'significant features of an author's personal stylistics' (Hori, 2004, p. 205). He also suggests that it is sometimes possible to correct critical intuition. Esther's language is usually seen as 'simple, plain and matter-of-fact' (Hori, 2004, p. 206) but her first-person narration shows a greater level of collocational creativity than the third-person narrator. Thus narratorial style is differentiated by Dickens by these stylistic differences. Hori concludes

corpus stylistics provides the benefits and strengths of comprehensive data, objective evidence and a more nuanced understanding of an author's use of language. Nonetheless, there are also pitfalls, particularly in the case of researchers who may automatically apply quantitative searches for distinctive linguistic features without reading literary works closely and sensitively. (Hori, 2004, p. 207)

Despite this last qualification, Hori's approach does seem to fulfil the goal of identifying stylistic features which a reader would be unlikely to locate, and to add to our appreciation of Dickens's work.

## 2.10 Combining corpus stylistic methods

As corpus stylistics matures, many of its practitioners are beginning to combine different approaches, both as a way of testing their efficacy and of exploring different texts. This process has been aided by the development of Wmatrix<sup>13</sup> by Paul Rayson at Lancaster University. Wmatrix allows web access to the CLAWS annotation tool, which is a part-of-speech tagger, as well as USAS, which is a

---

<sup>13</sup> Wmatrix software was developed by Paul Rayson of the Computing Department at Lancaster University. (Rayson, 2009)

semantic tagger. The analysis can be done on a user's own text, which they upload to the system. Results are generated 'at the word level to see *keyword* differences, at the POS level to see *key word class* differences, or at the semantic level to see *key concept* differences' (Ho, 2011, p. 118). Comparisons can be made with other large corpora, or another corpus provided by the user. Walker (2010, p. 367) points out that this is still fundamentally a lexical approach, with larger grammatical or semantic groups being identified through matching tags at the word level.

Using *Pride and Prejudice* as a text for exploration, Mahlberg and Smith (2010) analyse the vocabulary of the novel and illustrate a series of corpus stylistic methodologies. They begin with a simple keyword comparison between Austen's novel and the works of 18 other 19<sup>th</sup> century novelists, explaining that 'keywords are those words that are relatively more frequent in [*Pride and Prejudice*] than in the reference corpus' (Mahlberg & Smith, 2010, p. 251). The result is a list which is dominated by the names of characters, and then by personal pronouns and nouns denoting family relationships, such as sister. Mahlberg and Smith emphasise that this can only be a starting point, which might trigger an exploration of a particular word or a grouping of some keywords semantically.

The second method is the generation of a concordance around the node word *civility*. The concordance shows how this term is repeated, how it collocates with possessive pronouns as a way of characterising a particular person in the novel, and the positive and negative adverbs which qualify the term. This links critical literature on Austen's writing with the quantitative data produced by computer analysis. Mahlberg and Smith emphasise that, while corpus methods have the great advantage of investigating long texts as a whole, a second analytic step of this kind is essential.

Finally they discuss intratextual analysis, where the language of different parts of *Pride and Prejudice* is compared, rather than an external reference corpus being used as a comparator. This requires annotation of the text, and the available automatic annotation is limited, relying on punctuation marks of different kinds. However it does allow for 'suspensions' (described in section 2.9.2) to be identified, and generally to separate speech from non-speech. The suspensions are sorted into functional groups and links are found to the theme of civility and the expression of emotion. A high occurrence of suspensions, and their placement within the novel, illustrates their function. 'Suspensions are places that help to picture the synchronicity of speech and body language' (Mahlberg & Smith, 2010, p. 465). Coming together in linked associations and in a key scene in a novel, suspensions may add to the effect of such a scene for the reader. The 'crucial scene' would be identified by a reader without the help of a computer but the reoccurrence of suspensions might not be, so that foregrounding is achieved and the effect created without the cause being consciously identified.

Mahlberg and Smith then turn to a new approach, where they begin with existing literary criticism on *Pride and Prejudice*, which they can use to guide their exploration. They use the work of Korte (1997), who discusses body language in the novel and particularly the importance of eyes. Using eyes as a node word, with other words of similar meaning, a concordance is generated which confirms Korte's non-quantitative analysis. 'As Korte shows, descriptions of eye movement make it possible to give insights into characters' feelings that cannot explicitly be expressed' (Mahlberg & Smith, 2010, p. 460). However, the word eyes is actually less frequent in *Pride and Prejudice* than in the reference corpus, and therefore would not emerge in a keyword search. Mahlberg and Smith comment:

[t]his underlines that corpus methods need to be complemented by qualitative analysis. Although keywords can be a useful starting point, even a thorough analysis of all keywords cannot be taken as pointing to everything that is thematically relevant. (Mahlberg & Smith, 2010, p. 460)

Mahlberg's work on Dickens, in this paper and the research described in section 2.9.2, resembles my own project insofar as it is concerned with a prolific and renowned 19<sup>th</sup> century author and uses a corpus compiled and annotated from that author's works. However, our goals, methods of analysis and level of language for study are different; Mahlberg's analysis is based on the lexis of Dickens' novels whereas I am concerned with the syntax used by Henry James. However, I share Mahlberg's appreciation of the role of corpus analysis in making comparisons between long texts. Mahlberg and Smith illustrate the way in which 'corpus tools can guide the way into a text' (Mahlberg & Smith, 2010, p. 466), which resembles my analysis reported in Chapter 6.

Walker (2010) focuses on the semantic capabilities of the Wmatrix tool for his analysis of Julian Barnes's *Talking It Over*.<sup>14</sup> He is particularly interested in the narrators, pointing out that the novel is very unusual in having nine first-person narrators; he focuses on the three main ones: Oliver, Stuart and Gillian. His method involves comparing the words of one narrator with another, so that it is necessary for him to make separate text files for each one, which are then analysed by Wmatrix. This results in lists of keywords, key parts-of-speech and key semantic concepts for each narrator. He points out that, because of this division of the novel, the parts of his corpus are small and therefore statistical tests are less reliable; he has adjusted his statistical cut-off points accordingly.

The most salient key concept in the character Stuart's narration is called DISCONTENT and, along with the other key concepts, is considered by Walker to fit well into Stuart's circumstances in the novel. However, he sees the identification of

---

<sup>14</sup> Dan McIntyre also combines the use of Wmatrix with a semantic analysis in his work on the dialogue of the film *Reservoir Dogs* (McIntyre, 2010).

key concepts as only the first step. An examination of the words actually used to identify this key concept not only allows for the correction of any errors of allocation by the computer program, but also suggests words for more detailed analysis. This process brings Walker to focus on the lexeme *disappoint*, which he finds appears mainly in two clusters in the novel. Walker's discussion of these two clusters allows him to identify and discuss Stuart's preoccupation with, and changing attitude to, disappointment, both others' disappointment in him and the fear of being disappointed himself. Walker repeats this procedure with the other two main narrators. Walker concludes that

Wmatrix can be useful in stylistic analysis, providing an objective and replicable method of investigation of the narrators of a novel...It produced empirical data that supported my intuitions, as well as those of various critics...Importantly, in many instances, the insights that this "semantically guided" closer reading provided would have been missed by individual or keyword analysis alone. (Walker, 2010, p. 386)

However, he did find at times that Wmatrix was inaccurate in its categorizations, and suggests possible changes to the semantic categories it provides. Overall, the expansion of the corpus stylistics toolkit with more sophisticated tools of this kind is proving helpful in allowing corpus methods to truly enhance stylistic analysis.

Ho's book (2011) on the two texts of John Fowles's *The Magus* uses a number of different analytic tools to analyse the differences between the two versions of the novel and the effect of those differences. The two different versions were created by Fowles's own revision of the first published version of his novel, which was then released in a second edition. After establishing that the two texts can be differentiated by quantitative analysis, Ho, like Walker (2010) and McIntyre (2010), uses Wmatrix to analyse them, dividing the novel into different sections. Key semantic concepts are identified and Ho discusses both those which are used more in one version compared to the other and those which are used less. For instance, she discovers that the concept MOVEMENT is underused in the newer version, and links this to the suggestion that it 'does not focus on the story actions/events but on the (re)presentation of Nicholas's internal possible worlds' (Ho, 2011, p. 140). Ho goes on to utilise WordSmith Tools, generating a concordance which leads her to focus particularly on simile and metaphor. She discusses various metaphors used by Fowles, such as LIFE IS A PURPOSEFUL JOURNEY and LIFE IS A PLAY. From this exercise she concludes that 'the original conceptual metaphors ... are further elaborated and extended in Fowles's revision' (Ho, 2011, p. 186).

In a review of Ho's book, Wynne (2012) critiques its contribution to corpus stylistics. He argues that, as Ho does not use pre-existing language corpora nor compile a corpus of her own, her method is not corpus stylistic. Rather it is a quantitative analysis of the two versions of a novel. However, both McIntyre (2010)



and Walker (2010) use a process of internal comparison within a text in a process which is very plainly corpus stylistic analysis, so that the use of a large natural language corpus does not seem to be essential for inclusion in this field. Despite Wynne's claim, while much of Ho's work focuses on differences between the two versions of *The Magus*, she does also discuss the stylistic function of, for instance, key semantic concepts which she has identified using a computer. Her 'relevant reference corpus' is, in each case, the alternative version of the text. Ho's research is also relevant to the contrast between internal and external deviation discussed in section 1.2.1. The changes she finds in the revised version of *The Magus* are external to the new text but are arguably all part of the same work. Her findings fall on the borderline between internal and external deviation. Therefore I would argue that her work does fall within the remit of corpus stylistics.

The corpus stylistic approach can be combined with other stylistic approaches and themes. As this is not my approach, some studies are just mentioned here as part of the overall corpus stylistic framework into which my own project falls. For example, Toolan (2009) uses corpus stylistic methods (among others) to work within the framework of narratology.

Cognitive stylistics has emerged as a major field within stylistics. Semino and Culpeper explain that it encompasses

the kind of explicit, rigorous and detailed linguistic analysis of literary texts that is typical of the stylistics tradition with a systematic and theoretically informed consideration of the cognitive structures and processes that underlie the production and reception of language. (Semino & Culpeper, 2002, p. IX)

Thus a concern with textual analysis is combined in cognitive stylistics with an interest in how readers construe meaning from those texts. Work in this field is informed by theories from cognitive linguistics and cognitive psychology, and includes a number of different approaches, including metaphor, text world and mind style theories.

Examples are Heywood, Semino and Short (2002), Busse (2011) and O'Halloran (2007a).

## 2.11 Corpus stylistics and syntax

I take my definition of syntax from Aarts – 'the part of grammar that concerns itself with the structure of sentences' (Aarts, 2008, p. 4). Very little corpus stylistics has focused on syntax. Undoubtedly this is partly due to the limitations of suitable software which will parse automatically and accurately, and the resulting lack of fully parsed corpora available for both research and comparison purposes. Without accurate automated assistance, creating a fully parsed corpus of a literary work which a researcher may wish to study remains a laborious process. Where some

syntactic annotation is available and made use of, it is usually only at the level of parts of speech, be it an early project like that of Williams (1970) or relatively recent work, for example Stubbs's paper on Conrad (2005). Milic (1972) notably tried to go further but few researchers tackle language beyond the level of the phrase. Toolan (2009) is a notable exception, but he looks at the sentence, or matrix clause, semantically rather than syntactically. The development of Coh-Metrix at the University of Memphis (McNamara, et al., 2005) includes some syntactic analysis, but it does not provide a parsed corpus. Coh-Metrix is described in more detail in section 5.2.1.

Biber suggests that 'it is easier to notice the stylistic importance of word choice, while grammatical characteristics are much less salient' (Biber, 2011, p. 19) but admits that, in his own work with Conrad, 'registers and individual styles are distinguished by the use of pervasive grammatical features' (ibid. p.19).

Hoover's 1999 book, unusually, does include some limited discussion of syntax. He revisits Halliday's (1971) seminal paper on Golding's *The Inheritors*, and Fish's (1980) critique of that paper (discussed in section 1.2.2), using corpus stylistics to provide a new way of looking at *The Inheritors*. Hoover uses a number of different corpora including LOB (see footnote 11), a collection of contemporary novels held electronically, and so available for concordancing (the Ten Novel Corpus), and a specially made corpus comprising the first 50,000 words of 30 novels (the Novel Corpus).

With these tools, Hoover is able to compare vocabulary concerning body parts in *The Inheritors* and his Novel Corpus. His results show not only frequency but also syntactic function – whether a word, or string of words, is subject or object, for instance. He finds that

there are several times as many examples [of sentences in which body parts act as agents, as subjects of verbs of perception and mental process, and as subjects of intransitive verbs] in *The Inheritors* as in all of the novels of the Ten Novel Corpus and the Novel Corpus combined. (Hoover, 1999, p. 77)

Hoover interprets this as being 'largely responsible for the animism of the novel, an animism that is less a function of literary interpretation than it is of normal language processing' (Hoover, 1999, p. 77). However, Hoover qualifies this, saying that the linguistic evidence would not be enough to draw such a conclusion, which is also informed by other elements of the content of the novel.

Hoover concludes from his examination of the lexis of *The Inheritors* that it is unusual, although he cannot entirely explain why. When he uses reference corpora to compare the most common words in *The Inheritors*, he finds that Golding uses some words much more frequently and others far less. In other words, *The Inheritors* does not resemble other contemporary novels in its use of lexis. There are also unusual syntactic features, such as the use of 'short simple

sentences, mainly in the simple past tense' and 'a small, concentrated, peculiarly distributed vocabulary of short words' (Hoover, 1999, p. 155). Hoover tries to test his findings by putting these structures into other novels or removing them from *The Inheritors* and concludes that 'a constellation of textual features can [...] help to illuminate how the fictional world of a novel can be created and sustained' (Hoover, 1999, p. 168).

Mullender (2010) discusses the complexity of Shakespeare's late style by analysing his use of the word *which*, in response to

a widespread perception amongst generations of literary critics that what makes [the late plays] distinctive has something to do with language, that something changes in the linguistic fabric of the Shakespeare canon around the time of their writing. (Mullender, 2010, p. 11)

Yet she finds that 'there seem to be no attempts to describe exactly, empirically, what the linguistic changes are in the late group' (Mullender, 2010, p. 12). Much of the discussion of this question she considers to be 'impressionistic' (ibid. p. 15) and unquantified. There is quantitative work on the language of Shakespeare's late plays in the field of authorship attribution, but this has its own limitations, with the texts examined only partially, and a lack of clear explanations of the method used or the resulting data.

Mullender embarks on a more objective study by using WordSmith Tools and comparing the six late plays with a reference corpus of all the earlier plays. The keyword list includes many names of characters and places as well as words which describe the subject matter of the play. The word *which* is also very key and Mullender uses *which* as the centre of a detailed study of its use, making a concordance list which is then analysed manually, both by Mullender herself and two independent analysts. Interrogative and relative uses of *which* are distinguished, as well as restrictive and non-restrictive clauses and the length of the relative clause. Mullender also quantifies the number of words and of phrases between the pronoun *which* and its antecedent, and between *which* and the related finite verb which follows it. This has a resemblance to my quantification of the notion *delay* which is discussed in section 4.4.3. Mullender finds a distinct change in Shakespeare's use of *which*:

In the late plays, there is a very pronounced development (with the highest statistical significance) in Shakespeare's syntactic habits in *which* clauses, from the use of long phrasal interventions to the right, to interventions that are shorter and clausal, particularly containing two-word, non-finite clauses. It seems possible that this is a manifestation of what a number of literary critics have seen as characteristic of the late plays, namely a more syntactically and semantically condensed, as well as increasingly complex, style. (Mullender, 2010, pp. 205-6)

Further to this finding, the function of this change is considered. Mullender rejects a thesis that it is related to characterisation and suggests rather that it reflects Shakespeare's interest in the prose writing of the period.

Finally, Mullender discusses both the usefulness and limitations of WordSmith and other similar software, pointing out that keyness and frequency do not identify all features which may add to syntactic complexity. It is also unclear how much an infrequent but marked feature may be considered to affect literary style; as an example she cites 'six cases of large, multi-word phenomena such as double pushdown constructions<sup>15</sup> in a full length play such as *The Winter's Tale*' (Mullender, 2010, p. 234). She considers manual analysis as an adjunct to automated processing to have been useful because it adds to the accuracy of the results but found that it was very onerous. This study of Shakespeare's language has shown the possibilities of studying the syntax of literary texts using corpus stylistic methods, but also the limitations of a non-parsed corpus. The study of one pronoun, while indicative of a change in style, is still extremely limited, as Mullender acknowledges.

Boyne (2009) explores foregrounding using the notion and occurrence of 'deviant sentences' in two post-apocalyptic novels: *Riddley Walker* by Russell Hoban (1980) and *The Road* by Cormac McCarthy (2006). He defines deviant sentences as sentences which 'do not conform to what is considered prescriptively to be "grammatically correct"' (Boyne, 2009, p. 2), noting, however, that there is an oral quality in both novels and that what may be foregrounded as deviant will vary from speech to writing. He also emphasises the importance of the context in which a sentence is found. Boyne describes the syntactic deviance he finds in these novels at the clause level as either 'underdevelopment', characterised by the lack of a verb or a main clause, or 'overdevelopment', which features 'two or more main clauses blended without any conjunctions' (Boyne, 2009, p. 4). He considers sentences with these constructions to be foregrounded and analyses their stylistic effect. Both novels have similar syntactically deviant sentences, although McCarthy's overdeveloped sentences display polysyndeton rather than the lack of conjunctions Boyne identifies in Hoban's text. Boyne considers that, despite this latter structural difference, the stylistic effect of both the novelists' unusual sentence structure is likely to be the same.

---

<sup>15</sup> Mullender offers an example of a double pushdown from *The Two Noble Kinsmen*:

It is true, and I will give you comfort.  
To give your dead lords graves, the which to do  
Must make some work with Creon.

She explains that the word 'which' here is the object of 'to do' and also the subject of 'make some work with Creon' (Mullender, 2010, p. 130).

Boyne analyses three passages from each novel, the longest passage containing nine orthographic sentences and the shortest containing only two. In his discussion of the *Riddley Walker* passages, he considers that

the interspersion of complete sentences with deviant ones helps to foreground the latter while also establishing a sense of grammatical order: large scale deviance is not reader-friendly, and Hoban cannot afford more reader alienation than he is already risking with his orthographic deviance. (Boyne, 2009, p. 9)

Boyne considers the deviance of the syntax in *Riddley Walker* to mirror the collapse of order represented in the novel. He considers that the reader becomes relatively comfortable with the overall unfamiliar structure of the text, which is not random but connotes a new order in the far future world of the novel. Nevertheless, enough of a foregrounding effect remains to maintain the distance between our world and that of this dystopic novel. Boyne also sees the deviant syntactic structure as mimicking speech, as this is a first-person narration with some direct address to the reader.

Boyne's analysis of *The Road* also reveals a pattern within the deviant syntax:

a pattern that will recur throughout the novel: a complete sentence using a verb denoting action, followed by a deviant sentence (verbless or participial) that is descriptive of the scene evoked by the governing sentence. (Boyne, 2009, p. 14)

Where polysyndeton is employed, it often connotes repeated or habitual actions. Again, Boyne considers the syntactic structure used by McCarthy to be a stylistic device:

verbless and participial structures help to project a world that is frozen in time ... and polysyndetic structures depict with aching monotony the endless cycles of events required for the man and the boy simply to eke out their survival. (Boyne, 2009, p. 15)

Boyne explains that he is not claiming that syntactic deviance creates all the stylistic effect of these novels, but that it works together with orthography and lexis to do so.

Although Boyne does not differentiate the two, there are two kinds of deviance at work in these two novels. Syntactically deviant sentences display external deviance as the sentences differ from the internalised 'correct' grammar of the reader and are foregrounded on that account. Internally, Boyne identifies a contrast between those sentences which are grammatically correct in both novels (however unusual their lexis and punctuation may be) and those which are not. He gives the ratio between these two types of orthographic sentence as around 50:50. (Boyne eschews the term 'ungrammatical':

I stop short of calling them ungrammatical, since, as I will show, deviant sentences play an important contextual role in flowing prose that enables them to contribute both to meaning and to the projection of a fictional world in a much less limiting sense than prescriptive grammar alone might allow. (Boyne, 2009, p. 2)

While this study reveals a link between style and syntax, the amount of data which is analysed is very small – only 16 sentences from *Ridley Walker* and 15 sentences from *The Road* – and there is no indication that the extracts were chosen randomly. However, the claims which Boyne makes are for the novels as a whole, although the extracts are unlikely to be representative and the novels may not be stylistically consistent.

## 2.12 Conclusion

This survey of corpus stylistics has revealed some long-term trends which are mirrored in my project, and also a gap which this dissertation fills. Echoing stylistics in general, I share the concern which runs through many studies of marrying unavoidable (and not at all undesirable) subjective literary analysis with the use of quantitative data. This tension can be seen in the contrast between Carroll's and Milic's work in the early years of the discipline. There is a natural objectivity in the authorship attribution work of scholars like Burrows, and in the analysis of types of language by Biber and his colleagues. Many corpus stylisticians (Hoover, Semino and Short, Hori, Stubbs) discuss the need for literary critical input to make use of the data which corpus stylistics can reveal, and that is the position which I support. Such data can also help to fulfil the stylistic goal of replicability

Undoubtedly the history of corpus stylistics has been shaped by the restriction imposed by the software available. While nearly every writer describes the danger of being overly seduced by computer power into simply producing lists of statistics with no critical meaning, what can be done relies on what the computer can do. This in turn relies on what software programmers decide to produce, no doubt influenced by researchers' work and requests. The result is a complex mix of methods, shown, for example, by the work of Mahlberg as well as the fusing of corpus methods with other approaches, as evidenced in the work of Walker.

The theme of deviation, bringing about foregrounding effects, also runs through much of corpus stylistics. The comparison of a text with an appropriate reference corpus allows for the identification of external deviation while corpora are compiled from texts to be studied so that internal deviation can be found. It is then possible to analyse the results to establish whether there is foregrounding and to

what stylistic end. This theme links to my own use of the *HJPC*, described in Chapters 5 and 6.<sup>16</sup>

There are obvious practical reasons why corpus stylistics focusing on syntax is extremely rare, despite the studies discussed in section 2.11. However, there seems to be no principled reason for excluding syntax; on the contrary, both Milic and Culpeper argue for the importance of syntax in stylistic analysis. Leech and Short include a 'checklist of linguistic and stylistic categories' (Leech & Short, 2007, pp. 76-77) in their influential description and exemplification of stylistic analysis which includes syntax in various different ways. Where syntax is analysed, often the sample used is very small, as in Boyne's discussion of deviant syntax and Mullender's analysis of Shakespeare's use of a pronoun.

For my own project, applying corpus stylistics to James's style involved creating a parsed corpus. Although the search results of the corpus are objectively verifiable, there are subjective elements to the project which I will highlight in Chapter 4. In Chapter 7 I will discuss the extent to which my analysis, detailed in Chapter 5, adds new insights on James's style and where it confirms the work of previous literary critics. To this end, in the next chapter I will describe the views of James and others on his style.

---

<sup>16</sup> See, for example, section 5.2.5

# Chapter 3 Henry James and Style

---

In this chapter I examine writing by James and others which deals with James's ideas on style and their application in his writing. Literary criticism about James is an enormous industry which I can only sample, so I will just deal with some of the more relevant work here. I will also describe James's understanding of literary style in order to highlight the likely impact of his theoretical stance on his own novels. James's writing on style is represented in section 3.1, while in section 3.2 I give some examples of criticism which discuss James's style, finishing with those that use quantitative methods and are therefore closer to my own methodology. Section 3.3 discusses the two novels which I study in this dissertation, *Washington Square* and *The Golden Bowl*. When I refer to the novels as a whole, their names are used in full. *WS* and *GB* are used to refer to that part of the *Henry James Parsed Corpus (HJPC)* which is compiled from each novel respectively. Again I can only touch on the criticism which has been written on these works, especially in the case of *The Golden Bowl*. I outline the themes which have been important in the study of the two books and focus on work which is related to my quantitative study. As my research question attempts to identify the differences between the early style represented in *Washington Square* and the late style represented in *The Golden Bowl*, I discuss those critics who raise this question.

## 3.1 James's literary criticism

This section will review James's discussion of his own work and of that of other authors insofar as this reveals James's literary theories. Henry James was intensely interested in literary style in the broad sense of how novels should best be written. He was a prolific literary critic, as well as a commentator on his own work in his remarkable Prefaces to the New York edition of his novels. In his discussion of his own style, however, he does not dwell on the details of the language he uses in his writing; rather, he is concerned with the role of literature as a part of artistic work in general, and how it might best achieve its purposes. Moreover, he did not feel that style and content could be separated more than notionally. He wrote to the translator of his work (as quoted by Welleck): 'I feel that in a literary work of the least complexity the very form and texture are the substance itself and that the flesh is indetachable from the bones!'<sup>17</sup> (Welleck, 1958, p. 316)

In 'The Art of Fiction', originally published in 1884 in *Longman's Magazine* (James, 1948 [Original work published 1884]), James responded to a pamphlet of

---

<sup>17</sup> From *The Selected Letters of Henry James* (Edel, 1955, p. 171)



the same name by Walter Besant (1884), using the opportunity to expound his own theory of the novel. James welcomed the debate as previously, he said, the general belief had been that 'a novel is novel, as a pudding is a pudding, and that our only business with it could be to swallow it' (James, 1948 [Original work published 1884], pp. 3-4). In James's view, a novel must represent life, which it would do by presenting 'the author's vision' (James, 1948 [Original work published 1884], p. 10). As Blackmur explains, James's attitude was that 'the subject of art was life, or more particularly someone's apprehension of the experience of it' (Blackmur, 1935, p. xv). James's conception of the purpose of a novel is thus at one remove, in that it is not life itself, but life as seen by the author, which is represented. The portrayal is then removed further when James gives the reader not the author's own impressions but those impressions filtered through the consciousness and perceptions of the characters, very often a deliberately-inserted intelligent observer. Even when the narrator speaks, we are unsure whether this is actually James or a fictional narrator with their own personality which may colour their interpretation of the situation. So even before considering the considerable difficulties of the late style, there is complexity in James's view of the nature of literature. For James, a novel's value would depend essentially on the sensitivity of the author.

[I]t is an immense sensibility, a kind of spider-web of the finest silken threads suspended in the chamber of consciousness, and catching every airborne particle in its tissue. It is the very atmosphere of the mind. (James, 1948 [Original work published 1884], pp. 10-11)

With such refinement of perception, an author would be able to portray his own experience but also imaginatively portray other lives, using small impressions as seeds for the imagination. The worth of a piece of literature lay not in the subject matter but in how it was depicted. While literature should be true to life, 'the deepest quality of a work of art will always be the quality of the mind of the producer' (James, 1948 [Original work published 1884], p. 21).

For James, his method was best achieved in *The Ambassadors*. He explains his use of 'reflectors' in his Preface to *The Wings of the Dove*, telling us that there is not only one in a novel. (In modern stylistic terms, a 'reflector' would be a focalizer: the character through whom the reader learns about the action of the novel.) For James, they can 'work [...] in arranged alternation' (James, 1935, p. 301). He particularly likes the term 'reflector' because it punningly expresses not only their relaying of the plot through their eyes but also that they are 'burnished [...] by the intelligence, the curiosity, the passion, the force of the moment' (James, 1935, p. 300). For example, James explains in *The Golden Bowl's* Preface that the novel unfolds in the consciousness of first the Prince and then the Princess. Both narrator and reader are also observers, each at a further remove. In Welleck's

view, James's reflectors could also be a type, such as innocence in Maisie, although they must also be a fully integrated part of the novel. These ideas form part of James's experimental and radical approach to writing and were influential subsequently. Schwarz claims that 'by shifting focus from external action to the drama of consciousness, James foreshadowed interior monologue and stream of consciousness' (Schwarz, 1993, p. 47). However, despite James's stated dislike of an intrusive narrator, and criticism of Trollope for breaking the illusion of the novel, which he called 'little slaps at credulity' (James, 1948 [Original work published 1884], p. 59), even in *The Ambassadors* his narrator does intrude. In the much-discussed first paragraph the narrator comments 'on the principle I have just mentioned as operating' (James, 1998 [Original work published 1903], p. 1), showing that James still wanted the reader to be aware of the narrator (or author?) while suspending disbelief enough to be immersed in the start of Strether's story. The inconsistencies in James's use of reflectors are discussed further in section 3.3.2 with reference to *The Golden Bowl*.

In James's view it would not be desirable to attempt to formulate rigid rules as to how a novel should be written. Each author had to develop their own methodology, which would be intensely personal, but which should develop from their sensitive observation of life, 'catching the very note and trick, the strange irregular rhythm of life' (James, 1948 [Original work published 1884], p. 16). However, a novel must also have an essential unity and, though unpleasant aspects of life might be portrayed, a moral quality. It was also essential that a novel be enjoyed, and James praised Trollope for this quality but deprecated his huge productivity, feeling that he sacrificed quality for quantity. On the other hand, he was strongly influenced by Balzac, particularly in his early work, and admired the scale of Balzac's portrayal of the *comédie humaine*.

In Blackmur's (1935) collected edition of James's Prefaces, he discusses the view they reveal of James's opinion of his own novels. Unlike James, Blackmur comments on the novels' language, considering the late style 'an idiosyncrasy so powerful, so overweening, that to many it seemed a stultifying vice, or at least an inexcusable heresy' (Blackmur, 1935, p. xiii). He recognises the difficulty this may cause for the reader, both in the novels themselves and in the Prefaces. He attributes the late style to James's intelligence and to his extreme efforts to communicate his immensely subtle ideas to the reader. While James did not acknowledge the difficulty of his style, he did comment on his characters, who his readers might struggle to understand. He also wished to be economical in his work, condemning the 'loose baggy monsters' of Victorian fiction (James, 1935, p. 84). What James strove to achieve was a harmony of form and style, which Schwarz summarises as 'a text that, when its technique and subject matter fuse into a whole, will appeal to the reader' (Schwarz, 1993, p. 46).

In addition to his admiration for Balzac's (and George Eliot's) 'realistic' portrayal of life, James was also influenced by Turgenev and Hawthorne, who he felt represented respectively a facility with the form of the novel more usually seen in French novelists and the vital element of character and ideals. While he condemned authors who made plain the intervention of the author/narrator, breaking the illusion of a novelistic reality, first-person narration was also undesirable. James was striving for a more objective view, which would require a narratorial distance.

Welleck's review of 'Henry James's Literary Theory and Criticism' (Welleck, 1958) brings together a wider range of James's criticism. He does not consider the Prefaces to be basically critical in content, but rather as 'primarily reminiscences and commentaries' (Welleck, 1958). Welleck notes James's emphasis on the importance of enjoyment and appreciation for critics as much as any reader, and his notion that the critic's job is to look at treatment rather than subject matter, which is the author's own domain. James distinguished the novel, which is part of art and therefore a reflection of, and moral judgement on, an aspect of life, from romance, which is a story told mainly for fun, though not without value. Drama was a central influence on James's novel-writing and his literary criticism. For him, the novel should emulate drama; in Welleck's words, 'the novelist should let dialogue grow, to compose by scenes rather than by summary panoramic narration and description' (Welleck, 1958, p. 309). Welleck sees James as retreating somewhat from a too-close emulation of drama, striving for dramatic unity and economy but realising that description must also be used. James's use of the terms 'scene' and 'picture' is related, and will be discussed in section 3.2.3.

### **3.2 Literary criticism on James**

The field of Jamesian literary criticism forms a vast industry with a long history, tracking the developments in literary criticism since James's time. Tanner estimated in 1968 that 'every year brings forth around forty articles, three or four books, and an unknown number of dissertations on [James's] work' (Tanner, 1968, p. 11); no doubt the output is much higher today. It is possible to read works viewing James's oeuvre through a multiplicity of different lenses. In the bibliography given in the Oxford Encyclopedia of British Literature's entry on James, works cited feature Henry James and race, psychology, contrasts between American and Europe, anti-Semitism, sexuality including queer studies, modernity, feminism and post-colonialism (Haralson, 2006, pp. 132-3). It is clearly impossible to do justice in this dissertation to such a range. In this section I will simply give some examples of criticism closely related to the question of Henry James's literary style, culminating

in studies relating directly to his use of language, including some which use a quantitative and, in some cases, a corpus stylistic methodology.

Much of this criticism turns around the question of the 'difficult' late style, analysing its characteristics and discussing whether the difficulty is justified for the effect achieved. Gifford expresses the feelings of many readers:

[The three late novels] are notorious for their difficulty ... The notation is almost excessively fine, the issues often appear tenuous, the atmosphere has been pumped 'gaspingly dry'. Readers who delighted in the pictorial brilliancy of his earlier work and its neatness of style, must now grope in a world where for all the animation of James's figurative speech both meaning and action often hang in suspense; they must give unremitting attention to a new kind of discourse – the passional language of disembodied intelligences. (Gifford, 1983, p. 126)

In contrast Beebe, in an article first published in 1954, defends 'the much-disputed later style' which he considers to be 'the necessary medium of a highly individualized, subjective consciousness' (Beebe, 1968, p. 71). In 1961 Hopkins offered a more balanced view:

James's late style can be called either mannered or Mannerist, depending on whether one views his hesitations and qualifications, his inversions and twisting of syntax, his mingling of literary with colloquial language as an artifice masking an emptiness of content or as a mode of expression reflecting his painstaking effort to communicate with precision refinements of feeling and thought. (Hopkins, 1968, p. 113)

In addition to this range of views there is a variety of critical approaches to James's work, which will be described in the rest of this section.

### 3.2.1 Close reading of James

In F. R. Leavis's 1948 delineation of 'the great tradition' of English literature, he includes James as one of its members, though Leavis is also critical of James's style. Leavis considers *The Portrait of a Lady* to be 'one of the great novels in the language' (Leavis, 1972 [Original work published 1948], p. 147). He describes James as having a cult-like status, particularly around the late novels, but disapproves of *The Ambassadors*, which he felt might even show signs of James becoming senile. He preferred *The Bostonians*, considering James successful in his symbolism and seriousness, his productivity and his inclusion of a moral and psychological element in his work. However, Leavis is critical of the late James, both the novels and Prefaces. He deprecated the complex late style, citing an Edith Wharton anecdote to suggest that James had even begun to talk in an extremely mannered way and was not in control of the style, or perhaps even aware of it. This was not his only criticism; he felt that James did not have a rounded idea of his characters, with gaps in the narrative where it is not clear what they have been doing. Also James's elaborate imagery displeases him: 'We are conscious in these figures more of analysis, demonstration, and comment than of the realizing

imagination and the play of poetic perception' (Leavis, 1972 [Original work published 1948], p. 193). Leavis is accusing James of an intrusive narratorial presence, which is the very effect James says he is striving to avoid. However, Leavis still concludes, based on the novels which he admires, that James's achievement is a great one.

In an essay originally published in 1951, Raleigh discusses the development of James's style and particularly how James portrays his characters, who he considers to be rather similar throughout James's works.

If there is a change in psychological portrayal, it is one of extension rather than depth; that is, the characters and their reactions to situations are the same, but James has deepened and enriched their *effect* on the reader by all the resources of the late style, and the greater part of the power of the late style results from the fact that the concepts of consciousness which in the early novels were only vaguely implicit in the characters and their situations have now become explicit in the style. (Raleigh's italics) (Raleigh, 1968, p. 59)

Raleigh explains that James effects this change by showing the working of his characters' minds rather than describing and analysing them, so that the reader experiences the character directly, though in a controlled way. Despite the directness of this method, Raleigh admits that readers may still be puzzled 'by the peculiar splendour of the conscious lives and by the subtly shifting relationships of the characters in *The Wings of the Dove* and *The Golden Bowl*' (Raleigh, 1968, p. 59). Raleigh contrasts the early style, represented by *The American*, the middle style, represented by *The Princess Casamassima*, and *The Golden Bowl*'s late style, particularly focusing on the hero or heroine of each novel. He finds the heroes/heroines very similar but the style to have changed from the 'straightforward' and 'simple style' (Raleigh, 1968, p. 59) of *The American* to the more complex presentation of Hyacinth in *The Princess Casamassima*. With Hyacinth

the psychological process, while basically the same, is beginning to acquire power and depth. James had not yet arrived at the full-fledged dramatic method; so much about Hyacinth is described rather than presented, but the description is becoming fuller, more elaborate, and more concrete. (Raleigh, 1968, p. 60)

This direct presentation, or 'dramatization', becomes more marked, Raleigh claims, as the book progresses and 'the metaphors of this dramatization begin to dazzle in the late manner' (Raleigh, 1968, p. 60). Raleigh's suggestion that the style of James's middle period is closer to his later novels than the early ones is echoed in Hoover's (2007) quantitative analyses described in section 3.2.6.

Raleigh sees the central theme of *The Golden Bowl* to be abstract: the problem which exists in the relationships between the four main protagonists, which can never be plainly stated and is symbolised by the flawed golden bowl. The action

of the novel is the growth in Maggie's knowledge of what is happening, and in the realisation by some of the other characters that she knows. While Raleigh's approach is very different from my quantitative study, his article characterises the three stages of James's style, pointing out differences in James's presentation of his characters which are expressed in the changing syntax which I will describe in Chapter 5.

Mizener (1966), writing an Introduction to *A Reader's Guide to Henry James*, also acknowledges the difficulty of James's style and, like Leavis, connects it with James's life in general, while expressing more approval of its efficacy.

In the style of both his life and his work, James became throughout his long career steadily more mannered. By the end of it, his work - like the architecture of Vanbrugh or the lyrics of Hopkins - had a splendour so "high"[...] as to appear at first glance beyond ordinary comprehension. Only when one becomes familiar with it does one see the ironic, colloquial ease that controls it, and understand that, if James deliberately developed his famous manner because it was the best means available to him for saying what he had to, he was fully aware that artifice, though necessary to eloquence, is also absurd. (Mizener, 1966, p. 8)

Mizener justifies the complexity by explaining that what James was attempting to express was unusually complicated and that he was striving for a very fine quality of expression. He identifies such features as ellipsis and metaphor as part of James's 'strange dialogue' (Mizener, 1966, p. 8) but says that, at its best, it is beautiful. He accepts, however, that not everyone comes to an appreciation of James. Mizener connects James's interest in the inner life of his characters with his American roots and the interest in transcendentalist philosophy he shared with his father and brother. He explains: 'The necessarily subjective and personal apprehension of the quality of experience was as genuine a part of reality for James as was its social performance' (Mizener, 1966, p. 14). It is this sensibility linked with the European tradition of the novel of manners which Mizener takes to explain James's complex and controversial style

which had somehow to represent states of the consciousness and the accompanying uncertainties about objective reality in terms of the novel of manners. The cost was considerable; perhaps it was too great. (Mizener, 1966, p. 14)

### 3.2.2 James and Modernism

Many have made a connection between James's style and the later development of Modernist writers, and particularly 'stream of consciousness' techniques. In an analysis dealing with James and gender, using Lacanian methods, Boren (1989) makes comparisons between James, Proust and Joyce. Taylor (2003) sees James as part of the transition from the realism of the 19<sup>th</sup> century to early 20<sup>th</sup> century modernism, seeing the late style as James's means of showing the workings of his protagonists' minds. Hocks (1993) draws the same comparison between James and

Modernist authors but emphasises the difference between James's goals and the later development of a 'stream of consciousness'.

The stream-of-consciousness method, named and explained, ironically, by James's brother William in *The Principles of Psychology*, was, as interpreted and practiced by the moderns, more messy and wasteful than the writing of James, who sought not to transcribe the stream of thought but to dramatize it, largely through metaphor and adapted soliloquy (Hocks, 1993, p. 5).

Leech and Short's (1981) detailed literary linguistic examination of James's 'The Pupil' will be examined in section 3.2.6, but it is worth noting here that they too link the style of the story with James's desire to depict the workings of Pemberton's mind, and suggest that this supersedes a direct telling of the events of his story.

### 3.2.3 Drama, picture and point of view

Frantz Blackall (1993) gives a particularly clear explanation of James's desire to convey his plots not by explanation but by illustration.

James formulated the analogy that would dominate his conception of novelistic form ever after. The well-made novel, like the well-made play, should be scenic in design, fashioned in structural blocks and built up according to a carefully plotted dramatic scenario, with preparation, crisis, and denouement. (Frantz Blackall, 1993, p. 148)

James himself explains that *The Wings of the Dove* and *The Ambassadors* divide themselves 'into the parts that prepare, that tend in fact to over-prepare, for scenes, and the parts, or otherwise into the scenes, that justify and crown the preparation' (James, 1935, pp. 322-3). The power of this is that 'from an equal play of such oppositions the book gathers an intensity that fairly adds to the dramatic' (James, 1935, p. 326). Lubbock (1921) interprets the Jamesian key terms, 'picture' and 'drama':

Picture and drama, to him, represented the twofold manner towards which he tended in his last novels, composed as they are in a regular alternation of dramatic dialogue and pictorial description. (Lubbock, 1921, pp. 110-111)

For Lubbock, these are useful ideas in James's discussion of other novelists as well as his own work, and as a way of describing the relationship of the reader, narrator/author and character. In a dramatic scene, in either a novel or in a stage play, the reader/playgoer relates directly to the character, making their own judgement without authorial guidance. The role of the writer may be inferred but is not made evident. In a 'picture', the narrator intervenes, describing the scene for the reader. Lubbock acknowledges that 'drama' cannot be total in a novel; some narratorial intervention is essential but the idea is still dominant. Lubbock particularly notes how in *The Ambassadors* the action is viewed entirely through Strether's consciousness: 'His mind is the mirror of the scene beyond it, and the other people in the book exist only in relation to him' (Lubbock, 1921, p. 146). *The*

*Ambassadors* is, therefore, the great exemplar of the pictorial method but with a reflector rather than a narrator, and the reader is able to watch Strether's development directly as the action influences his thinking and emotions. Suspense in the storytelling is achieved by Strether himself not knowing how his feelings will change (which is the real story of the novel rather than the events). Scenic (or dramatic) episodes are not entirely foregone, because we see Strether's encounters with Maria Gostrey and there we hear the dialogue with no intervening thought; this gives the reader some necessary distance from Strether.

For Fowler (1993), this innovation is part of the contrast between James's early and late novels, the latter having fewer characters but, she considers, greater depth. While the late novels use the technique of focalizers,<sup>18</sup> Fowler considers that they resemble James himself very closely, in that they require so much insight to tell their stories. However, I suggest that, while this may be true of Strether, it cannot be said of the Prince in *The Golden Bowl* who so misjudges Maggie's intelligence and strength. While themes link early and late novels, the treatment is different as James becomes interested in showing the reaction of his characters rather than the plot itself; in this sense, the use of one or more focalizers may be considered to bring the reader closer to the action where the focalizer is, as in *The Golden Bowl*, one of the main protagonists. Fowler also identifies the late style as part of this project to convey the finest possible feelings and complex perceptions, along with James's elaborate metaphors. 'The characters of the late fiction thus accrue more meanings for us than can easily be delineated, taking on the three-dimensionality and complexity of actual people' (Fowler, 1993, p. 181).

The use of one or more focalizers is also often discussed in terms of 'point of view'. Chatman (1998) is critical of James's way of using the term (although James described the concept, rather than actually using the more modern word 'focalizer'), while acknowledging his seminal work in the development of narrative theory.

Originally (in James, Lubbock, and Booth) it meant something like "position--whether spatial, perceptual, moral, or whatever--from which a story is presented." The term made no distinction between the points of view of characters, narrator, or author (real or implied). (Chatman, 1998, p. 378)

Chatman insists that it is impossible to ignore the narrator in James's work, who sometimes actually intrudes in the first person, however much an omniscient viewpoint may be minimised.

---

<sup>18</sup> Leech and Short (2007) define a reflector (or focaliser [sic], which they say is now the most common term) as 'the person whose point of view is reflected' in a fictional narrative (Leech & Short, 2007, p. 140). However, they warn that that focaliser may change frequently, from sentence to sentence or even within a sentence.



### 3.2.4 James's metaphor and imagery

James's use of imagery and particularly the extended metaphors of the late novels are much commented upon. Gale's *The Caught Image* is particularly relevant here in that it uses quantitative methods. However, Gale gives no details of his method for counting imagery. Presumably they were all noted by hand in this pre-computer study. Unfortunately this means that his work is not replicable by other stylisticians. He does explain that an image comprises 'the tenor (or literal aspect of the comparison) and ... the vehicle (or figurative aspect)' (Gale, 1964, p. 6). He gives figures for the density of imagery in James's novels, as shown in Table 3-1.

**Table 3-1 Image density in James's fiction**

Works of fiction	Words	Figures of speech	Density/1,000 words
135	4,189,800	16,902	4

(Gale, 1964, p. 8) [Table my own]

Gale explains that the density of James's use of imagery varies remarkably little through his career, though there is some identifiable increase in use from the 1880's to the era of the late novels in the early years of the twentieth century. There is also an unexplained correlation with the setting of the novel: more imagery is associated with novels set in England than in France, in Italy rather than America. Gale also points out that some of James's images are extremely extended, so that in the New York edition there is roughly an image on every page. As Gale acknowledges, these raw numbers are not very helpful so he makes some comparisons with other American fiction writers. Hawthorne's *The Scarlet Letter* apparently has a somewhat higher metaphorical density, with Melville being similar to James, *The Great Gatsby* somewhat lower and *Huckleberry Finn* lower still. Gale goes on to categorise James's imagery according to its content. He notes that

the largest single category of similes and metaphors in the fiction of James [...] [is] that of art. Nearly two thousand separate figures - more than a tenth of the total - concern art in one of more of its forms. [...] The relative frequency of the art figures increases from the 1860's to the 1870's and then drops steadily but only slightly. (Gale, 1964, p. 102)

This is an unsurprising finding given James's deep interest in art and his self-identification as an artist.

In his statistical appendices, Gale gives figures for the two novels featured in this dissertation. *Washington Square* is described as having 143 images in 62,200 words, which is a ratio of 2.3 images per 1,000 words. *The Golden Bowl* is richer in imagery with 10,902 images in 192,200 words, a ratio of 5.7 images per 1,000

words. Gale also provides a table of image density by decade, slightly adapted here, with the decade for my two novels shown in bold:

**Table 3-2 Image density per 1,000 words by decade**

1860's	1870's	1880's	1890's	1900's	1910's
3.5	3.9	<b>3.1</b>	4.2	<b>5.4</b>	4.8

(Gale, 1964, p. 254)[Table adapted from Gale's]

The contrast in richness of imagery seems to apply not just to *Washington Square* as compared to *The Golden Bowl*, but to their whole decades, which are at the extremes of James's range in this area. The reason for this is unclear. Importantly for my purposes, these data suggest that, in terms of metaphoric density at least, the two novels may not be representative of the early and late styles in general, although the three great novels of the late period (*The Wings of the Dove* 1902, *The Ambassadors* 1903 and *The Golden Bowl* 1904) were all written in the 1900's. Gale also shows that long novels have the highest density of imagery, followed by short stories, short novels and long short stories in descending order. Although this seems quite random, it again polarises *Washington Square* and *The Golden Bowl*, as does his claim that revised works have more imagery than unrevised works; *The Golden Bowl* text used for the *HJPC* is from the (revised) New York edition.

Yeazell suggests that 'in the late Henry James, dangerous knowledge tends to be mediated by way of powerful, extravagant, and deeply disturbing metaphors' (Yeazell, 1976, pp. 39-40).<sup>19</sup> Cohn finds that this applies particularly in the first chapter of the second book of *The Golden Bowl* (otherwise Chapter 25), where the Princess becomes the focalizer. She points out that there is no dialogue in that chapter, which I will show to be unusual (see Figure 6-6). Maggie introduces the vivid image of the pagoda to represent her situation and there are many other shorter metaphors in a chapter focused inside Maggie's mind. Cohn sees Chapter 25 as a particularly difficult section of *The Golden Bowl*:

[i]magery accumulates more profusely and concentrates more exclusively on what happens within his fictional figure; the tension between authorial and figural narrative situations is optimal; an incomparably complex time structure conveys simultaneously what the protagonist experiences and what she remembers. (Cohn, 2001, pp. 6-7)

Cohn considers chapters 25 and 26 of *The Golden Bowl* to be 'one of James's most daring experiments in the internalization of fictional action' (Cohn, 2001, p. 7).

---

<sup>19</sup> Previously Raleigh had made a similar suggestion (1968, p. 60).

### 3.2.5 James's language

Although they represent a small proportion of the vast range of Jamesian criticism, some writers do examine James's language in detail, concentrating on the characteristics of the late style. Levin (1986), for example, makes some detailed observations on James's syntax and its application in his introduction to *The Ambassadors*:

[s]entences can be prolonged but seldom periodic, with relatively few subordinate clauses and a good many parentheses. Word order is rhythmically varied by inversions and expletives. Formality in structure is relieved by a plasticity in style, which searches not so much for the *mot juste* as for the gradual approximation, the continuous modification, the qualifying nuance. Since the work was dictated, it sounds vocal (note the contractions),<sup>20</sup> though with James's highly literary voice. (Levin, 1986, pp. 27-28).

It is unclear what comparator Levin is using in claiming that there are 'relatively few subordinate clauses'; I count the actual frequency of subordinate clauses in *WS* and *GB* in Chapter 5 as part of a wider measure of syntactic complexity (see section 5.2.2).

Springer, also commenting on *The Ambassadors*, points out the variability of James's late style. She notes that at the end of the book the sentences are very short, equating sentences 'full of breaks, of semiotic music, of qualifications' (Springer, 1993, p. 278) with strong emotion. This accords with my association of short, simple sentences and strong emotion, described in section 6.5.1.4. Springer praises James's role as 'a developing linguistic revolutionary' whose language has an 'expansive strangeness' (Springer, 1993, p. 278) which has its purpose in showing the depth of the mind of the focalizer.

Menikoff's 1971 paper exemplifies a detailed focus on a particular aspect of James's prose. He associates James's late style with 'the attempt to simulate the process of the mind, the manner in which an individual apprehends or perceives an idea - and to engage the reader in that process' (Menikoff, 1971, p. 436). This is similar to James's own explanations of his goals, and of commentators such as Blackmur, described in section 3.1, although with more emphasis on the role of the reader. Menikoff explores the idea of the representation of mental processes by looking at James's use of the third person singular pronoun, for example James's use of 'she' in his story 'Julia Bride', which starts with that word without any further introduction to the focalizer. Although there is an implied narrator, Menikoff points out that we soon feel that we are seeing through 'her' eyes, so that the method is very close to the first person narrative which James rejects. He notes that 'she' is

---

<sup>20</sup> However, Virginia Llewellyn Smith points out in her Notes on the Text to *GB* that many of the negative contractions were introduced when James revised *GB* for the New York edition, and so were a reaction to reading rather than a result of dictation (Llewellyn Smith, 2009, p. xxxiii).

often combined with verbs of mental action, and they are often used as a conditional so that the reader is drawn into 'her' thought process.

Cross also looks in some detail at James's language, having first noted, once more, the difficulty it causes readers as they share feelings of uneasiness and disorientation with his characters. She locates the problem specifically in his sentences, with 'their constant flicker and spill of meaning' (Cross, 1993, p. 1) and agrees that some of the features of his style may be due to his habit of dictating his work. She emphasises the radicalism and innovation of James's prose:

James's is a syntax that takes grammar to the limit and outplays its codes; it is radical in that it usurps word order and complacencies of grammar, revolutionising the way meaning could be disseminated over a text. (Cross, 1993, p. 2)

Cross notices a difference in style between James's early and late work, though without giving quantified data for her analysis. She discusses the frequent use of compound sentences in his early work, and uses the word 'doubling' to describe sentences where a proposition is expanded or contradicted in second and subsequent main clauses. This shifts the focus, Cross explains, from the object(s) described to a number of contrasting meanings they may carry. On the subject of the late style, Cross uses the early sentences of *The Ambassadors* as an example and notes how James uses long clausal and phrasal constructions for single grammatical functions within the sentence. While compound sentences are a feature of the earlier work, later in his career she suggests that the parallels are expanded but also varied so that the symmetry is broken. A further characteristic which makes for difficulties in interpretation is the way that the meaning of a sentence may be carried not by the main clause but by a dependent clause, such as in the example below:

- 1) His "peak in Darien" was the sudden hour that had transformed his life, the hour of his perceiving with a mute inward gasp akin to the low moan of apprehensive passion that a world was left him to conquer and that he might conquer it if he tried. (James, 2000 [Original work published 1904], p. 82)

This sentence of Adam Verver's free indirect thought only comes to the point in the two conjoined dependent clauses which are direct objects of the verb 'perceiving', from which they are separated by 13 words. Cross particularly emphasises the way that James's prose remains within the rules of grammar and uses syntax to gain his effects:

If James takes grammar to the limit by relaxing its barriers, expanding its rules of syntax in lexical selections and patterns that are extravagantly marginal [...] none the less, it is important to add, *James still relies on the logic of grammar to carry his sentences through* (Cross's italics) (Cross, 1993, p. 23).

Although the grammar is technically correct, difficulty is also sometimes created by vague reference, and words which build up special meanings, such as the word 'wonderful' in *The Ambassadors*. This forces the reader to consider the paragraph as a whole, rather than the individual sentence they are reading, in order to work out to whom personal pronouns are referring. The multiplicity of references using abstract nouns adds to the effect, resulting finally in a text which Cross considers 'baroque' (Cross, 1993, p. 30).

### 3.2.6 Quantitative studies of James's style

Although it is a more unusual approach, there are a number of studies of Henry James's style which use quantitative methods. In particular, a series of critics across more than 40 years, often reacting to each other, have focused on the style of *The Ambassadors*. Though their approaches differ, to some extent they support each other's findings as well as echoing ideas discussed in non-quantitative studies.

Long before there was any question of using a computer or corpus methods, Vernon Lee in 1923 responded to a letter in *The Times* suggesting that it might be useful to apply statistical tests to literature to examine writers' style. She reports her findings for a number of authors, including Henry James. On page 127 of *The Ambassadors* she counted 500 words of which 137 were nouns and pronouns, 71 were verbs and 48 were adjectives and adverbs. Then, taking the first sentence which is not dialogue, she attempted a closer study. She found that there were many personal pronouns, which she considered gave a more personal effect than the use of nouns. Like many later commentators, she noted James's unusual use of pronouns: 'a sort of personification. There is, at all events, an extraordinary circling round these pronouns' (Lee, 1992 [Original work published 1923], p. 243).

Short (1946) also sets out to examine James's sentence structure, using the second chapter of the second book of *The Ambassadors*. He intentionally omits 'simple and direct sentences', focussing instead on 'the rangy, convoluted sentences that bear so unmistakably the hallmark of James' (Short, 1946, p. 71). Short assesses the length of James's sentences by comparison with Dr Johnson's *The Rambler* (though that dates from 1750-52), and comes to the conclusion that James's sentences are not unusually long. This surprising result points to the importance of a careful choice of reference texts; James's sentences may not be longer than Johnson's but may still be unusually long if Johnson's have the same exceptional proportions.<sup>21</sup> Short also points out that some of James's sentences consist of two or more constructions which could stand alone as sentences, joined with punctuation. Short calculates that in the 196 sentences in his sample, the

---

<sup>21</sup> Short is relying on Wimsatt (1941) who holds that Johnson's sentences 'are not unduly long as compared with various standard English writings' (Short, 1946, p. 72).

average sentence length is 35.3 words. When the 44 'loosely connected' sentences are dismantled, the average word count falls to 25.3 words (Short, 1946, pp. 72-73). (In my corpus the average sentence length in *The Golden Bowl* is 29.6 words but Short particularly selected long sentences for his sample.)

The finality, the crystallization, that ordinary sentence order and signs defining relationship bestow upon prose has been skilfully foregone in favor of other values. In these peculiar sentences, facts remain tentative, intentions fluid, and conclusions evanescent. (Short, 1946, p. 74)

This effect is heightened by the peculiarities of James's word order, which again Short relates directly to James's intentions. For example, he cites a sentence in *The Golden Bowl*, spoken by the Prince -

"How can I not feel more than anything else how they adore together my boy?" (James, 2000 [Original work published 1904], p. 177)

- and suggests that 'together' is so placed to emphasise it, showing how strongly the Prince feels about the close relationship between Maggie and her father, and hence how he feels justified in his relationship with Charlotte. Short concedes that arranging word order to achieve a particular effect is not uncommon, but claims that James goes far beyond the usual, creating an effect closer to poetry than prose.

Short links James's 'misordered' sentences with his frequent use of parentheses, feeling their import is to mimic the psychological state of the characters, although he admits that this is not limited to such characters as are confused, restricted or whatever. Instead he suggests that James's characters find themselves in situations where they have to rethink their ideas and beliefs. James's style represents this new and disorienting situation, and the reader is forced to adapt alongside the characters, sharing in their struggle. Nevertheless Short regards James's prose as cohesive:

[n]ew units tend to contain elements of orientation with a great deal, if not all, that has gone before; the meaning expands in a process of accretion. With each new unit a fresh atom joins the ring of fluid, organic, suspended meaning. (Short, 1946, p. 80)

Short suggest that ambiguity is caused by James's use of pronouns without clear referents. It is also clear that sometimes James uses these stylistic devices without clear reasons, apparently mimicking the pattern set by those sentences in which such unusual structures have been meaningful. Short calls this 'stylization' (Short, 1946, p. 82), although without being critical of this unnecessarily mannered style. However, Short also points out that some passages in the late novels are quite straightforward. One example is the opening of *The Golden Bowl*. (This echoes my findings, summarised in section 5.5.) However, Short criticizes these islands of clarity as being out of step with the style of the novels as a whole, and

praises James's general conformity to consistent stylization. James's lack of differentiation of the voices of his different characters, with the possible exception of Charlotte Stant, is also defended; James is most interested in 'perception and moral nature' (Short, 1946, p. 84), not in human individuality, and he is being controlled by his stylization. This argument is less than convincing; surely the interest of James's plots lies, at least partly, in the contrasting natures of his characters.<sup>22</sup> Also, as Short admits, James used a similar style in his Prefaces, letters and conversation. To some extent at least, the late style reflects James's mannerisms, whether they were consciously adopted or inadvertent.

Watt references Short's work in his seminal 1962 essay on the first paragraph of *The Ambassadors*. He notes that in his sample the sentences average 41 words, which is only a little longer than Short's finding. However, unlike Short, he denies that there is very much unusual about James's style in terms of syntax, finding the main problem to be reference. Watt's acceptance of James's syntax as unremarkable is surprising. Although the paragraph is not enormously 'stylized', to use Short's term, and the syntax is entirely correct, there are certainly some of the unusual word order and heavy use of dependent clauses which other critics have noted. Watt lists much use of non-transitive verbs, abstract nouns, 'that', elegant variation and negatives as the predominant characteristics of the late style. He quantifies these elements, for example counting 14 passive/copulative/intransitive verbs contrasting with only six transitive verbs in the paragraph. The preponderance of abstract nouns as subjects of clauses justifies the idea that James's prose is abstract in general, though the figures cited of four abstract subjects against three concrete ones are less than overwhelming, and there is no comparison given to other writers. More problematically, when Watt compared these results with eight other paragraphs, he found that there was a great deal of variation. While the tendency to use non-transitive verbs and abstract nouns continued, the use of 'that' and negatives did not. This undermines Watt's implied claim that his description is of the late style in general, rather than just this paragraph.

Watt remarks on James's use of parenthesis and suggests that many of the oddities of phrasing and word order are for comic effect, so that the reader and narrator are amused together at Strether's discomfort with his feelings. He also examines the structure of the paragraph as a whole. It consists of six sentences which can be divided into three pairs, each focussing on a different aspect of Strether's delay and becoming increasingly complex, with the fifth sentence being very long and the sixth suddenly short, like the punch line to a funny story. This structure allows for some suspense even though we are being shown thoughts

---

<sup>22</sup> I compare the speech of the Prince and Charlotte in section 6.6.1.4.

rather than actions. Watt concludes that whether the late style is entirely a conscious choice or not, James used language in the service of his idea of what a novel should be concerned with, and that in the late style he produced 'a narrative texture as richly complicated and as highly organised as that of poetry' (Watt, 1960, p. 271).

David Lodge (1966) links details of the plot of *The Ambassadors* with specific linguistic features. As we have seen, the average length of sentences in *The Ambassadors* has been counted by Short as 35 words (based on Part 2 chapter 2) and by Watt as 41 words based on the opening paragraph. In the passage where Strether realises Chad and Madame de Vionnet have been deceiving him, Lodge counts the average sentence length at just over 37, changing to just under 33 from the point of Strether's realisation. Lodge explains that this last figure is made higher by one exceptional sentence where Strether is trying to assimilate his new knowledge. 'His actual apprehension of the odd, disturbing behaviour of the couple in the boat is appropriately conveyed in short, alert sentences' (Lodge, 2002, p. 213). This pattern of an exceptionally long sentence in a paragraph of more moderate sentences matches my findings summarised in section 5.5.

Lodge continues to compare this later paragraph in *The Ambassadors* with Watt's analysis of the opening sentences. He explains that Watt's list of characteristics matches the first half of his selection better than the second half. For example, the first six sentences of this later section match Watt's ratio of 14 non-transitive verbs to six transitive exactly. However, the last six sentences have 12 non-transitive and 10 transitive verbs. Lodge does not find heavy usage of 'that', or elegant variation, or abstract nouns, unlike Watt. Overall Lodge endorses Watt's interpretation of the purpose of the late style, linking Strether's troubled mind with the violation of 'the linguistic norms of the novel' (Lodge, 2002, pp. 215-6). However, as with Watt's paper, general conclusions are being drawn here from an extremely small sample to a very long novel, or even to a group of novels, without any quantitative evidence that they are homogenous.

Other critics do look at James's novels as a whole. Chatman's *The Later Style of Henry James* is perhaps the best known detailed description of the language James uses in the late novels. Chatman's approach is, to quite a large extent, quantitative. He begins his book with an examination of samples of some of James's novels, comparing novels from different stages of his career and other contemporary works.<sup>23</sup> His figures are summarised in Table 3-3 (with James's novels in order of publication). Chatman's figures are scattered through his book and described in different ways, rather than presented graphically. My table brings

---

<sup>23</sup> The other novels Chatman used are Joseph Conrad's *Typhoon* (1902), Samuel Butler's *The Way of All Flesh* (1903), George Gissing's *Veranilda* (1904), E. M. Forster's *Where Angels Fear to Tread* (1905).



his data together. Chatman looks at the use of nouns as grammatical subjects referring to human beings, tangible objects or intangible elements, a slight elaboration of Watt's approach. The samples show that there is a progression from the early to the late novels, with more subjects referring to human beings in the early novels, and more intangible reference in the later ones, with *The Portrait of a Lady* intermediate in both date and intangibility. The novels by other authors are most like *The American* in this respect. Chatman also includes some quantitative comparison of other elements. For example, he tries to see if there is more use of verbs of mental action in the late style, but the results are less easy to interpret. *The Ambassadors* inexplicably breaks the pattern of the other two late novels and *The American* is not very different from *The Golden Bowl*.

The table also shows a number of grammatical features which Chatman explores in developing his main theme of the abstraction and intangibility of James's vocabulary and their relative predominance in the later style. He finds some evidence of both change between James's early, middle and late usage, and the contrast between his style and that of his contemporaries. He introduces the idea of 'plenitude' – the great wealth of detail in the protagonist's consciousness which is conveyed by abstraction, ellipsis and complexity.

Table 3-3 Chatman's figures - collated from different parts of *The Later Style of Henry James* (Chatman, 1972)

	No. of sentences	Grammatical subject nouns - human	Grammatical subject nouns - tangible	Grammatical subject nouns - intangible	Verbs of mental action	<i>It</i> expletive as subject of main clause	as if	Main clauses beginning with expletive <i>there</i>
<b><i>The American</i> (1877)</b>	200	71%	6%	23%	221	Not quantified	Not quantified	1
<b><i>The Portrait of a Lady</i> (1881) Chapter 42</b>	200	56%	2%	42%	Not quantified	'half again as many instances' as <i>The American</i>	3	15
<b><i>The Wings of a Dove</i> (1902) Book IV Chapter 1</b>	200	55%	1%	44%	250	8%	More than <i>The Ambassadors</i>	Not quantified
<b><i>The Ambassadors</i> (1903) Chapter 5</b>	196	45%	6%	49%	165	Not quantified	More than <i>The Portrait of a Lady</i>	Not quantified
<b><i>The Golden Bowl</i> (1904) Book IV Chapter 2</b>	200	53%	1%	46%	228	Not quantified	6 (most)	2 or 3 times as many in late novels than other novelists
<b>Other novelists</b>	200	67%	15%	14%-22%	Approx. 50-85 in 2 novels	Conrad 5%, Butler 4%, Forster 4%, Gissing 1%	0	Not quantified

Though Chatman recognises James's heavy use of parenthesis, he feels that it is the combination of embedding and ellipsis which makes for the real difficulty, and describes how the reader has to 'hold open a mental slot' (Chatman, 1972, p. 126) while they wait for the ellipted item to be given, only to find that it never is. (This has parallels with my concept of *delay* described in section 4.4.3.) For Chatman, as for most other critics, the significance of this complex syntactic structure is that it is a way of conveying a 'character's consciousness as he struggles with his problems' (Chatman, 1972, p. 127).

Also taking a range of James's works, in 1975 Laitinen looked at James's use of four rhetorical figures – exclamation, rhetorical question, emphasis and hyperbole. He is interested in the reader's reaction to James's writing, and in particular how emotion is read and felt, and makes the assumption that these devices are likely to be 'the most effective in externalizing emotion and thus conveying it' (Laitinen, 1975, p. 7). He examines samples of 20,000 words in longer novels, 10,000 in shorter ones, choosing four novels<sup>24</sup> ranged across James's writing career (while excluding the earliest phase), but selecting the actual samples at random from the early part of each book. Laitinen focuses on particular characters within the novels, comparing their rates of use of the figures with other characters. For example, he finds that Isabel Archer has an unusually high frequency of exclamations, hyperbole and rhetorical questions within *The Portrait of a Lady*. He draws conclusions from this about her character, which he sees as naive, and about the range of emotions she expresses, reacting to her situation. Laitinen's conclusions are not always easy to justify; for example, Ralph Touchett also uses a great deal of exclamation, but in his case this is taken to exemplify his ironic wit.<sup>25</sup>

Although David Hoover's (2007) work on James concentrates only on lexis, it provides useful evidence of the development of James's style across his career, as well as providing an example of corpus stylistics as applied to James. This double approach of evaluating corpus techniques while also investigating Henry James's style mirrors the research questions of this dissertation. Corpus methods allow Hoover to use a large proportion of the words of a novel rather than a small sample which might not be representative of the style as a whole. This solves a problem mentioned by Leech and Short (2007, p. 306) – namely that the defining features of an author's style may only become detectable over a large quantity of text, and be a matter of quantity as well as quality. Following Burrows's (2005) methods which are more usually applied to authorship attribution studies (see section 2.7),

---

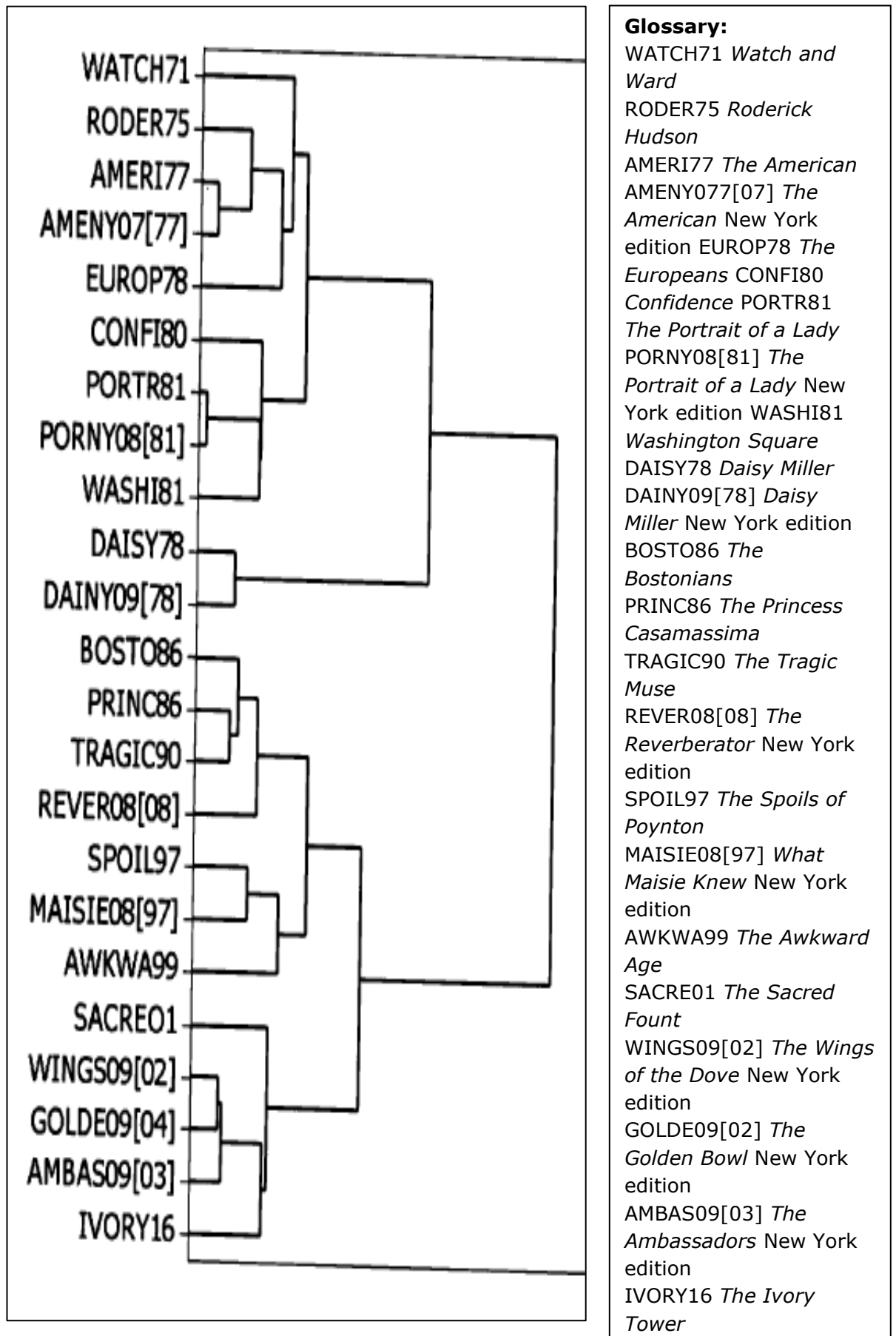
<sup>24</sup> *The Portrait of a Lady*, *The Aspern Papers*, *The Spoils of Poynton* and *The Ambassadors*.

<sup>25</sup> Laitinen's appendices includes graphs of the absolute frequencies of the four features in the four novels, cumulative totals and a breakdown of each feature for each character.

Hoover first establishes that he can distinguish James's work from that of his contemporaries, and then goes on to compare James's novels relative to their date. He creates a corpus of 71 novels by James and his contemporaries and uses both Cluster Analysis and Burrows's Delta technique. Both use the most frequent words in the texts as comparisons, and both prove effective in distinguishing the different novelists.

Hoover goes on to examine twenty of James's novels (all the major novels which are available as electronic texts) using a series of quantitative analyses. The Delta technique is almost completely successful in dividing the novels into three distinct phases of work: 1871-81, 1886-90 and 1897-1917. The gaps between these dates coincide with breaks in James's novel writing, and Hoover notes that a division of James's novels at least into early and late groups has been agreed by a number of critics (Hoover, 2007, p. 178). Cluster Analysis is even more successful, putting the novels into almost perfect chronological order, which suggests that James's style developed steadily throughout his career, though with three distinct phases. It shows the intermediate novels as being more similar to the beginning of the late phase than the early phase. The results of Hoover's cluster analysis of James's novels are shown in Figure 3-1, which reproduces Hoover's Figure 3. Hoover does not provide a glossary to his abbreviations of novel titles but they are given here to the right of the diagram. Finally, Hoover applies principal components analysis using the same word list as has been employed with the other techniques. Not only are the novels grouped into three separate periods, they also 'tend to appear in publication order throughout' (Hoover, 2007, p. 180). Where revised New York editions of early novels are included they appear on the graph a little closer to the later novels than their corresponding unrevised form, 'showing that James's extensive revisions make them more like the later novels without masking their dates of composition' (Hoover, 2007, p. 180). Hoover notes that not all authors show this kind of chronological development, though he comments that much more needs to be done to understand the development of authorial style. I have used Hoover's three periods to inform my choice of novels for analysis.

Figure 3-1 James's three phases (Hoover, 2007, p. 182)



Smit (1988) also uses quantitative methods in a book which is critical of many stylistic studies on James, suggesting that they do not make sufficiently explicit the reader's role or the extent to which a critical response is also a personal one. More importantly, some of the arguments are said to be weak or circular. For example, Fowler (1981) suggests that 'James' use of nominalizations expresses Strether's inactivity, 'but the major evidence we have that Strether is inactive is that James describes him with nominalizations' (Smit, 1988, p. 24). Smit also questions to what extent readers react similarly to James's style. While there is obviously some common ground, reactions will differ and Smit does not accept the idea that James's prose somehow educates his readers to appreciate his work as he wishes.

Smit identifies a list of features which other critics have considered typical of James. Quantitative analysis of such features, Smit suggests, must be treated with caution because items are chosen for counting subjectively, according to the interest and experience of the critic. Smit illustrates his argument by creating a list of Jamesian features as identified by critics and comparing them against passages he considers to lack the usual Jamesian style in *The Wings of the Dove*. He does not find the expected characteristics, but I suggest this is unsurprising; critics have usually accepted that there is variation even in the late novels. Smit then compares James's writing in various different genres, both published and unpublished, to see how varied it is, and finds that he uses shorter and more active sentences with few of the famous Jamesian mannerisms when he is writing letters and journals. Smit did not find a way to quantify the relative complexity or the degree of parenthesis which he found in James's sentences (issues which I will address in section 5.2). He is convinced that in private correspondence James's style is less complex, and also that dictation was not the cause of the more difficult and typical late style, nor that it was the result of some psychological state, as that, presumably, would have pervaded all his writing at the time. This leaves the problem of why James did write in this way in literary work. Smit's answer is:

his concern - we might even call it an obsession - for artistic form and his "sacred" task. When he was composing his public work, James used what I can only call an "aesthetic register". [...] he was demonstrating how the artist could delve into the depths, attune himself to subtleties, and project those subtleties in language. (Smit, 1988, pp. 78-9)

This process, Smit contends, was taken to an extreme because of James's isolation without an agent or editor or critics who would dare to question his method. Smit is also critical of the idea that James is attempting to portray the workings of the mind. Despite suggestions by other critics to the contrary, he maintains that the dominant point of view is that of the narrator. The mannerisms of the late style do not reflect the minds of the characters, who Smit claims all sound quite similar, but

the narrator himself. He still calls James a genius who conveys in a unique way the human experience of seeing the world through one's own eyes, but holds that the characteristic complexities are not part of that excellence but simply irrelevant, and not always, or necessarily, an asset. He even questions to what extent style, at least at the level of sentence structure, is within the conscious control of the writer. 'In many ways, the late style cannot be justified aesthetically; it is simply the way James wrote' (Smit, 1988, p. 93). This view is, of course, the opposite of James's, who felt his style was integral to his purpose, although he discussed style in terms of point of view, and direct or indirect portrayal of characters and events, rather than syntax.

Although there is little reference to it in his text, and despite his doubts about the objectiveness of any selection of text to be analysed, Smit undertook a quantitative analysis of some narrative passages in *The Wings of the Dove*, each of about 300 words, three of which were chosen at random and the last chosen because it seemed to be an exception to the expected style. Smit found that the average sentence length ranged from 34.5 to 53.3 words per sentence and that intangible nouns were used as sentence subject 30.8% to 63.6% of the time, while 23.1% to 30% of main clauses had verbs of mental action. (The rate of intangible noun subjects is substantially greater than Chatman's findings shown in Table 3-3.) Smit found it difficult to decide which parentheses were unusual, so he simply counted those which were indicated by punctuation in the text. In the random passages he counted 13, 11 and 13 parenthetical interruptions, respectively. He also provides figures contrasting James's published and private writing.

Smit's work cuts across much of the accepted wisdom on James's work. Perhaps most importantly, he contradicts James's own contention that he often works through a reflector and portrays the mental processes of his characters. Even if the narrator is more present than James suggests, Smit's position does not match the preponderance of focalization through a fictional character which characterizes much of James's work.

In another, more limited, study, Leech and Short examined James's 1891 short story 'The Pupil'. They describe various lexical features, and then justify James's sentence structure by saying that frequent parentheses provide a 'complex wholeness' (Leech & Short, 1981, p. 228). They find that James uses dependent clauses in a ratio of 3:1 with independent clauses – a level of use much greater than Leech and Short found in Conrad (and which does not accord with my findings illustrated in Figure 5-5). They note that *that*-clauses, with or without *that* actually appearing, are the dominant form, and relate this to psychological verbs, such as *know (that)*.

These apparent perversities of James's syntax become meaningful in the light of an appraisal of his particular concern with psychological realism: his

unremitting endeavour to pin down the psychological moment (Leech & Short, 1981, p. 102).

While the use of quantitative methods can be informative, and can produce some counter-intuitive results, such as James's apparently normal sentence length compared to his contemporaries and the identification of 'un-Jamesian' passages, it does not obviate the need for critical judgement. However, critical judgement does not create consensus. Perhaps a greater amount of carefully-analysed data would help to do so, though personal readings will always remain idiosyncratic. With the same texts to analyse, and reasonably similar analytic methods, critics can focus on different aspects and features, and come to different conclusions about James's intent and the novels' effects.

### **3.3 Style in *Washington Square* and *The Golden Bowl***

In this section I will discuss the two novels which are analysed in this dissertation, giving a short summary of their plots and publishing history, as well as a brief overview of some of the issues which literary critics have discussed when reading them. Again I will particularly focus on any quantitative studies. There is far less written on *Washington Square* than *The Golden Bowl*, as interest tends to concentrate on the late novels which most (though not all) consider to be the summit of James's achievement.

#### **3.3.1 *Washington Square***

*Washington Square* was first published in six monthly parts, in London's *Cornhill Magazine* from June to November 1880 and also in *Harper's Monthly* in New York from July to December of the same year. (The version I am using for my corpus was published by Macmillan in London in 1881.) The novel concerns Dr Sloper, a successful doctor in New York in the middle of the 19<sup>th</sup> century and his daughter, Catherine. Dr Sloper married a 'charming' heiress and has a successful medical practice. His son died aged three and his wife died shortly after giving birth to Catherine. Catherine is a disappointment to her father, as she is docile and obedient but neither pretty nor intelligent. Her chief companion is Dr Sloper's widowed sister, Mrs Penniman, a foolish, romantic and secretive woman. The action of the novel starts when Catherine is wooed by Morris Townsend, whom her father correctly suspects of being interested in her fortune.

In the central chapter of the novel, which I will examine in some detail in Chapter 6, Catherine confronts her father with her feelings about Morris Townsend. In her characteristically timid and submissive way, she attempts to obey her father while not immediately abandoning her hopes of marriage. Her father accuses Townsend of being mercenary and makes it clear that he will never leave an



inheritance to Catherine if she marries him. Faced with the horrifying idea of looking forward to her father's death, Catherine promises that if she does not marry Townsend in her father's lifetime, she will never do so after his death. This declaration is the turning point of her life; the novel ends with her still holding to her promise. She is not distracted by her father taking her on the Grand Tour during which father and daughter confront each other passionately while walking in the Alps, nor does she deviate after her father's death, when Morris Townsend tries once more to marry her for her money.

*Washington Square* was not one of James's favourite works: he failed to include it in the New York Edition, and there is, therefore, no Preface giving his own commentary on it. In a letter to his brother, William, he says that 'the only good thing in the story is the girl' (quoted in Shelston, 1984, p. 29). He mentions the origin of the story in his journal for February 21<sup>st</sup> 1879. He records that Mrs Kemble, a famous actress of the period, told him a very similar story concerning her brother's broken engagement to a rich but dull young woman. *Washington Square* is markedly different from the late novels in many ways, including the language James uses. Blackmur (Blackmur, 1959, p. 5) comments that 'the last thing he was writing in these books was the Henry James novel in any of the versions of it to which we are now accustomed'. He finds *Washington Square* to be a story to be read for the pleasure of it, but also notes the seeds of later James themes, such as the relationship of women to the conventions which attempt to constrain them.

Despite James's own misgivings, critics have often praised *Washington Square*. Leavis notes its resemblance to Balzac's *Eugénie Grandet*, but considers that it is an improvement on Balzac and an example of James's versatility (Leavis, 1972 [Original work published 1948], p. 161). Although *Washington Square* does not contrast American and European life and culture in James's trademark fashion, Buitenhuis (1970) discusses whether the relationship between Catherine and her father may be particularly American. He feels that the Doctor would like to love his daughter, and that the Grand Tour is his attempt to awaken her intelligence sufficiently so that he may do so, but the experiment fails. This point of view is unusual, with more critics finding the Doctor cold and unfeeling.

Maseychik's (1968) main criticism of *Washington Square* is that its characters are two-dimensional. 'The most striking thing about simple Catherine Sloper is her garish clothing' (Maseychik, 1968, p. 119). Maseychik finds even Dr Sloper to be shallow because of his unswerving decision to forbid Catherine's marriage to Townsend. The Doctor is 'as inevitable as a problem in geometry' while 'Morris himself is most like an inanimate object' (Maseychik, 1968, pp. 119-20). Travel changes nothing in any of the characters, in Maseychik's view, who sees Catherine's immobility at the end of the novel as a symptom of this lack of growth.

This view denies the growth in courage and strength of personality which, I would argue, Catherine undergoes, so that she surprises her father and ultimately surprises Townsend as well

Knights (1975), in contrast, is a fan of *Washington Square*, which he calls 'the small masterpiece'. He admires Catherine's growth into a person with real depth and the terror of the Alpine scene, which is unlike anything else in the novel. He particularly picks out the sudden realisation which the reader has at the very last sentence that the story is, after all, a tragedy. Knights, like Blackmur but unlike Buitenhuis, blames the Doctor for Catherine's difficulties, treating her with what he calls 'brutality' in his cold rejection (Knights, 1975, p. 16). This is carried through not only to the actual terms of the Doctor's will, but the language in which it is written. Le Fanu (1982) echoes Knights's praise for *Washington Square*, calling it 'the short, witty masterpiece' whose prose is 'so categorically brilliant, at all times so devastating in its accuracy and appropriateness' (Le Fanu, 1982, pp. vii-viii). Le Fanu goes so far as to question whether the development of James's late style was, in fact, an improvement. He compares *Washington Square* to *The Golden Bowl*, which he calls 'that great and flawed final novel of James's "major phase"' (Le Fanu, 1982, p. viii), suggesting that the clear and precise language of *Washington Square* does not, in fact, lack the subtlety which the late style is reputed to have developed, and that, had James revised *Washington Square* for the New York Edition, the result might have been damage rather than improvement to the novel.

Despite praise from some critics such as these, there is far less comment on *Washington Square* than on the late novels, especially detailed accounts. One exception is Lucas's (1972) examination of the novel. Again he is admiring, mentioning the novel's 'unmistakeable brilliance' and 'entirely justified self-confidence and sureness both in style and the handling of events' (Lucas, 1972, p. 36). Lucas also praises the comedy of the characters in *Washington Square*, and the way they are described to a large extent in dialogue. He discounts the idea that James mimicked *Eugénie Grandet* to a great extent, as he feels that *Eugénie Grandet* is focused on the heroine, whereas James is more interested in the social context.

Its drama depends on the range and subtlety of James's presentation of social relationships, with all that that implies of tone, habits of deference, poise, conscious civility, calculated decorum: all those elements, in short, in which certain lives are given definitive shape and to which they can become forfeit (Lucas, 1972, p. 38).

The society in which *Washington Square* is set is very much the New York of the 1840's which was, as Adrian Poole points out, the period of James's early childhood when his maternal grandmother and aunts were living in Washington Square (Poole, 2010, p. xiv). The customs and manners of that society were therefore well-

known to James, and Lucas holds that this is vital to the story, although he notes that others have disagreed. For him any criticism that the characters in *Washington Square* are caricatures is justified by the rigorous social mores of the setting, which constrain and shape ordinary lives. For example, Townsend is characterised, Lucas claims, even in his name, which labels him with the outsider, unfashionable status from which he is so anxious to escape. However James also stands outside this society, with his 'aloof distaste for the society he presents. There is a cool, detached and scalpel-sharp dissection of society at work in this novel, a kind of ironic ruthlessness' (Lucas, 1972, p. 42). There is comedy in Townsend's less-than-elegant behaviour when he is alone with Mrs Penniman, as well as in her assumed airs and graces. There is also irony in Dr Sloper's position: he is a self-made man himself, having married well and achieved business success. It is not difficult for him to recognise the same ambition in Townsend. However, Lucas is not unsympathetic to the Doctor, who he feels is trying to reveal himself to Catherine in the dramatic scene in the Alps and is angry that she cannot understand him.

The position of the narrator who does not overtly judge any of the characters is congruent with James's theory that 'the novelist should be an invisible narrator of his own tales' (Lucas, 1972, p. 48). Ultimately none of the characters is completely bad, and the relationship between Catherine and her father is complex and nuanced. For Lucas, it is the way the story is handled by James which gives it its value.

The brilliance of James's novel depends on the way in which its comic surface is played off against the tragic events, so that nothing strident is allowed to substitute for its unruffled and sure study of circumstance, of context, of concatenation of cause and effect. It is the ordinariness which is so extraordinary about *Washington Square*. (Lucas, 1972, p. 59)

For Bell (1975), *Washington Square* is much closer to melodrama with a cruel father, false lover and ostentatious aunt, although she also sees it as a critique of melodrama because James is self-conscious in his use of the form and interested in realism. For example, Dr Sloper is a gentleman but also a working man, a particularly American persona. Bell also sees Catherine as an American type.

Her American quality is the opposite of his: it meets experience by means of the promptings of a virgin nature, the expression of a human condition or national origin uninstructed by the past. (Bell, 1975, p. 21)

Bell sees *Washington Square* as pivotal in James's development as a writer, allowing him to experiment while preparing for what she sees as his first great novel, *The Portrait of a Lady*. *Washington Square* is a play between what is natural and unnatural, as well as an examination of intelligence and how it is used. The character of the narrator, Bell suggests, is particularly closely identified with James

himself, with his own narrative style, which is ironic and metaphorical. Lucas (1972) also discusses *Washington Square* in terms of melodrama, but argues that James moves beyond the archetypal characters recognised by Bell in the development of Dr Sloper's attitude and Catherine's maturity. He suggests that Dr Sloper and James/the narrator are very closely identified in their comic and ironic attitudes to Catherine and her aunt at the beginning of the novel. Later their positions separate as Dr Sloper is unable to accept that Catherine is a human being with real emotions and her own dignity. 'He [Sloper] can only comprehend the surface of her character because he wants to believe that the rest of it is a settled matter' (Poirier, 1984, p. 53). This superficiality is the essence of melodrama but it applies to Dr Sloper and not to the novel as a whole, because James's understanding of his characters is more nuanced and perceptive.

Le Fanu (1982) is also appreciative of the narrator of *Washington Square*; he emphasises the gentle wittiness of the narrator's language and, above all, the precision. For example, he cites Catherine's summary of her life: 'From Catherine's point of view the great facts of her career were that Morris Townsend had trifled with her affection, and that her father had broken its spring' (James, 1998 [Original work published 1880], p. 180). Le Fanu asks: 'Does James [...] ever formulate anything as cogently as this in his later, elaborate masterpieces?' (Le Fanu, 1982, p. x). Le Fanu sees *Washington Square* as centring on the development of Catherine's character, which is misunderstood by the Doctor, and actually everyone in the story including herself at the beginning of the book. Focusing on Catherine's own growing strength of purpose and resolution, *Washington Square* may be closer to the interiority of the late novels than Le Fanu has acknowledged. What Le Fanu admires is the nuanced portrayal of the characters, so that the Doctor is not, he argues, a melodramatic type but real, with genuine concerns for his daughter but a coldness at his heart which may be a reflection of an inner despair, and possibly religious disbelief. Le Fanu also writes approvingly of the comedy element in *Washington Square*, personified by Mrs Penniman, and also the tragedy she brings out. Le Fanu finds it especially bitter for Catherine that her fate is determined by Mrs Penniman's stupidity, and sees the Aunt's betrayal of her niece in transferring her allegiance to Townsend as 'the single most treacherous action that the novel sees fit to relate' (Le Fanu, 1982, p. xxi).

Those who have written in detail about *Washington Square* often prefer its simplicity to the later James novels, though some also see it as a preparatory exercise. The major disagreement among the critics is the extent which the characters in the novel are seen as rounded human beings and to what extent they are melodramatic types.

### 3.3.2 *The Golden Bowl*

James was much happier with *The Golden Bowl* than he was with *Washington Square*. Written in 1903 and published in New York in two volumes in late 1904, *The Golden Bowl* was then published in one volume in London, and never serialised like *Washington Square* and many of James's other novels. James then reissued it as volumes XXIII and XXIV of his New York Edition, although *The Golden Bowl* was not radically revised as it had been written so recently. It was his last complete novel and was not a commercial success at the time. The idea for *The Golden Bowl* had come to James many years before 'when he heard that a middle-aged American and his only daughter had simultaneously become engaged to be married' (Llewellyn Smith, 2009, p. vii). The idea that the two spouses might become entangled was James's own.

*The Golden Bowl* is often cited by critics as James's best novel, although others prefer *The Ambassadors*, or *The Portrait of a Lady*. Fowler, for instance, calls it 'in many regards his most brilliant [novel]' (Fowler, 1993, p. 198). Leavis is an exception, being no fan of the late style and disapproving of what he felt was an unduly sympathetic portrayal of the acquisitive Ververs. 'That in our feelings about the Ververs there would be any element of distaste Henry James [...] seems to have had no inkling' (Leavis, 1972 [Original work published 1948], p. 185). The novel has attracted an enormous amount of critical interest. A comparison of articles in *The Henry James Review* finds 74 results when searching for 'Washington Square' but 273 for 'The Golden Bowl'. The articles on *Washington Square* are quite wide-ranging, focussing on the novel's adaptation as a film and comparisons with other authors, but also on geographical locations, the social setting and discussions regarding the role of women. The articles on *The Golden Bowl*, however, are more heavily balanced towards ideas, themes, symbolism, literary theory, character and plot, although they also cover film, art and other authors. Clearly I cannot do justice to this range of criticism and here will simply touch on a few key issues to illuminate my own analysis of the language of the novels.

*The Golden Bowl* has a considerably more complex plot than *Washington Square*, centring around four main figures – Adam Verver and his daughter Maggie, who are wealthy American collectors and travellers, an impoverished Italian prince, Amerigo, and Charlotte Stant, who has previously been romantically involved with the Prince but parted from him because they were too poor to marry and maintain a princely lifestyle. Charlotte is now a close friend of Maggie's and at the opening of the book, the Prince has just become engaged to Maggie, but neither has told her of their previous relationship. Charlotte engineers a last day with Amerigo in which they shop for a wedding present and the Prince rejects a gold-covered crystal bowl because it is flawed. After the Prince and Maggie Verver are married, Maggie continues to be very close to her father and the Prince and Charlotte are thrown

together a great deal. Still unsuspecting, Maggie is concerned that her father may be lonely and decides that he should marry Charlotte. After the second marriage, Maggie and her father continue to spend most of their time together. In Chapter 22, Charlotte and the Prince finally consummate their adulterous relationship when they are left behind at the end of a country house weekend. It is this incident which begins to arouse Maggie's suspicions, which are then confirmed when she buys the golden bowl herself, and the Jewish vendor recognises a photograph of the Prince, remembering his pre-wedding shopping expedition with Charlotte. Maggie's knowledge is revealed, albeit obliquely, to the reader by means of a conversation between her and Mrs Assingham. Confronted with the flawed golden bowl as evidence of the relationship between Charlotte and Amerigo, Mrs Assingham smashes it, but the truth has been brought to light irrevocably. Maggie protects her father and her own marriage by contriving that her father should take Charlotte with him back to America without telling him of her discovery. This short plot summary does nothing to bring out the subtlety of James's story, which unfolds obliquely, with the reader and Maggie both struggling to discern the truth. Some matters never clarified, such as the true extent of Adam Verver's knowledge.

Fowler (1993) touches on the question of the late style with reference to *The Golden Bowl*, saying ambiguously that 'in it, the possibilities and limitations of language are tested to their breaking point' (p. 198). While she seems to mean this as praise, it does suggest difficulty, and she sees the language as much as a way of concealing knowledge as of revealing it. She also mentions James's pervasive European/American theme which here she says is taken to its most extreme, with the richest possible American and the most aristocratic possible European. However, her main interest is in the development of Maggie Verver's character, although seen abstractly as the type of Self and Other. In *The Golden Bowl* Maggie (and her father) are seemingly unable to relate to the Prince and Charlotte as other than acquisitions, but Maggie's awakening leads to a possibility of a real relationship within her marriage, although at enormous cost for her relationship with her father and Charlotte's entire future.

A far more critical view of the style of *The Golden Bowl* is taken by Springer (1993), who quotes views that it is an upper class or masculine display, arid and circular. However, she does not dismiss it entirely, suggesting that its ambiguities may be an intended reflection of Maggie's ambiguous and doubtful hopes for her future. 'Convoluting language [...] can also be seen to work for the effect of a given piece, as well as letting us know the depths of our implied author' (Springer, 1993, p. 279).

Ralph Norrman's detailed study of *The Golden Bowl* focuses particularly on the ambiguity of reference, which he explains as being James's deliberate working out of the elaborate relationships he has set up within the novel.<sup>26</sup> The two marriages, Amerigo and Charlotte's affair and Maggie's unhealthily close relationship with her father make for complications before the relationships between Maggie and Charlotte and between the two men are added:

Once the basic situation was given James needed some intensifying device to dramatize the successive stages in the development of the plot. A device that very naturally suggested itself was referential ambiguity. This ambiguity usually occurs in some form of misunderstanding in a dialogue. (Norrman, 1982, p. 9)

Norrman explains that although some of the ambiguities are left to mystify the reader, usually they are explained by some device such as an interruption and serve 'to remind the reader regularly of the combinations and relationships in *The Golden Bowl* (Norrman, 1982, p. 10). With the multiple relationships at the heart of the novel, the ambiguity is often in personal pronoun reference: when the women refer to 'he', they could mean either man, and when the men say 'she' they could mean either woman. The complexities of the possible reference of 'they' are accordingly even more multiplied, and 'you' could be singular or plural. For example, very early in the novel Norrman points out an ambiguous reference which foreshadows for the reader the existence of an important woman in the Prince's life who is not Maggie. The Prince is speaking to Mrs Assingham:

You know with what care I desire to proceed, taking everything into account and making no mistake that may possibly injure *her*. (James, 2000 [Original work published 1904], p. 23)

The italics are James's own. Mrs Assingham responds by asking who '*her*' refers too, the Prince replying that he means Maggie and her father but the reader is left wondering why Mrs Assingham might have doubted it. The question is resolved but, Norrman suggests, the drama heightened when Mrs Assingham reveals shortly afterwards that Charlotte, the Prince's former lover, is in London.

Yeazell (1976) highlights ambiguity of a different kind; she suggests that it is difficult for the reader to decide who knows about the Prince and Charlotte's affair, and also difficult to make moral judgements about the protagonists. James's method of narration through a focalizer who is also a character in the drama gives no guidance, and painful feelings are concealed, rather than revealed. 'The late style demands at every point we sense more than we are yet able to articulate; only gradually do we grow fully conscious of our own subliminal guesses' (Yeazell,

---

<sup>26</sup> Horne points out that ambiguity of reference is a 'late Jamesian practice' which can be seen in *The Ambassadors*, 'The Turn of the Screw' and 'The Figure in the Carpet'. 'Mysteries of reference are James's stock in trade' (Horne, 1997, p. 119).

1976, p. 35). This uncertainty as to moral judgement is seen in the varied views critics take of Maggie in particular. Wright (1968) explains that she is characterised by some as a witch and by others as a saint; his own view is that these are oversimplifications and that Maggie is a nuanced character by the end of her growth through the novel. Wright also conjures an Edenic myth for Adam and Maggie, with Maggie having to acquire knowledge and Lebowitz (1968) echoes this, although she seems to see Maggie and Amerigo, and then later Adam and Charlotte, as the primal couple. This is a little far-fetched but Lebowitz's linking of *The Golden Bowl* to a fairy tale is more plausible, with the Prince and Princess as the central couple of the tale and Fanny Assingham as a kind of fairy godmother. However, she points out, *The Golden Bowl* goes beyond the fairy tale as the Princess has to face reality before the somewhat ambiguous happy ending.

It seems James's use of pronouns is quantifiably greater in the late style,<sup>27</sup> and, in *The Golden Bowl* at least, this use of reference as part of the narrative technique may be the explanation. It follows that the reader's task is not so much to disambiguate the text as to be informed by all the possibilities that the ambiguity allows. Norrman traces a lessening of pronominal ambiguity as the Prince and Charlotte grow closer to consummating their affair, and so become more clearly a couple, an 'us', until they join erotically with a switch to the possessive pronoun 'their' (Norrman, 1982, p. 27). The working out of the pairs is paralleled by a change in pronominal use until Maggie is referring to Charlotte and her father as a couple as she consolidates her own relationship with her husband. The climax of the novel's closing is when the Prince uses the pronoun 'you' to mean, unambiguously and only, his wife.

The acknowledged ambiguity of *The Golden Bowl* is not always criticized; Dryden (2010) links ambiguity and metaphor but enjoys the difficulty, discussing it in almost Jamesian style (possibly an occupational hazard of Jamesian critics):

The delightfully complicated first chapter of Book 2 of *The Golden Bowl* - with the prolonged, hovering presence of the authorial consciousness, generating the circling temporal tiers and elaborate network of extended metaphors that at once contain and extend the simple actions represented in the chapter's thirteen pages, teasing and turning them this way and that - may be read as a sort of synecdoche of the entire novel. (Dryden, 2010, p. 112)

Dryden sees metaphors as inherently ambiguous, and links them, and James's 'complicated syntax', to 'the imp that seems to lurk within social institutions while still maintaining the fictions that the intentions and ideals, the rational forms by which we live, remain intact' (Dryden, 2010, p. 122). In other words, for Dryden

---

<sup>27</sup> Norrman (1982) cites Robert G. Johnson's dissertation, 'A study of the style of HJ's Late Novels' (1971) Bowling Green State University, Ohio.



the ambiguity of James's writing is the mirror he holds up to the ambiguous social world his characters inhabit, with improper actions hidden and unspoken to preserve the proprieties. The shattering of the golden bowl threatens the outward harmony of this world but it is promptly put back together again by Maggie, both literally and figuratively in her engineering of the geographical separation of the four protagonists into their married partnerships.

Norrman identifies another idiosyncrasy of the late style which he calls 'end-linking'. By this he means:

the linking of two linguistic elements such as phrases, sentences, periods or paragraphs end-to-beginning, usually through the repetition, at the beginning of the following element, of something at the end of the preceding. (Norrman, 1982, p. 66)

Norrman provides as an example parts of two sentences from *The Golden Bowl*:

...Mr Verver whose easy way with his millions had taxed to such small purpose, in the arrangements, the principle of *reciprocity*. The *reciprocity* with which ... (Norrman's italics) (Norrman, 1982, p. 68)

This device, he suggests, serves to give cohesion to James's writing but also to foreground the form of his prose over the content. The effect is intense, reflecting back on what has been said, rather than forward moving. 'Each unit in James's text is very much propped up against its linguistic neighbours rather than merely resting on its referent' (Norrman, 1982, p. 67). However, the result of repetition can be to increase ambiguity rather than to clarify, but in this case it is no accident.

James's aim was to make content important by presenting it through a medium made important. If the participants in a dialogue find it important to make out the surface-meaning of their partners' utterances this should also give stature to the reality behind that intensely elaborated surface. (Norrman, 1982, p. 67)

Norrman strongly defends James's skill exemplified in the late style, even if he accepts that it is arguable to what extent the multiple parentheses imitate actual speech. He characterises James's use of parenthesis as keeping all the movable elements as near as possible to the beginning of the sentence, with a definite closure, commonly a verb, in a sentence pattern which is reminiscent of German. Any check in the flow which this may cause is bridged by repetition, so that the forward movement is constant. The result is a text 'in which everything remains tentative' (Norrman, 1982, p. 71). Norrman does admit that James's constant use of anadiplosis<sup>28</sup> can be seen as mannered, and that it may be linked to his dictation of his books. It may make it difficult for the reader to follow the argument because

---

<sup>28</sup> Reduplication; a figure in Rhetoric, when the same word that ends one part of a verse or sentence, is repeated in that which follows [quotation 1656 from Thomas Blount *Glossographia*] (Oxford English Dictionary, 1991, p. 1538).

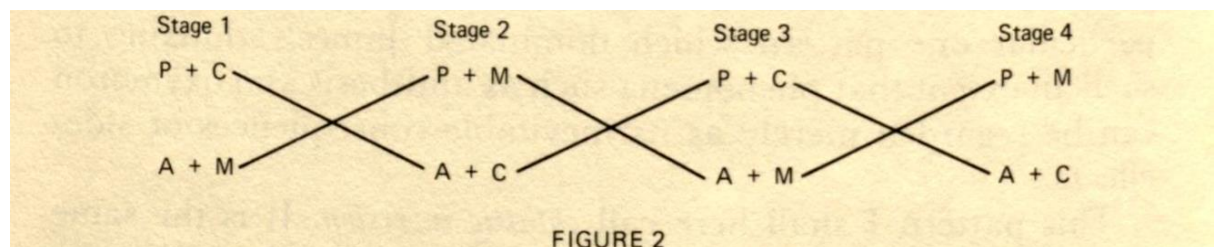
there is a constant backward reference to the preceding sentence. Certainly the late novels have to be read with great attention and without skipping. Synonymy and elaborate imagery add to repetition sometimes as banter between the protagonists but also to add to the intensity of a situation. Repetition is also used as affirmation: the second speaker agreeing with the first but not simply saying yes. Norrman suggests that this strengthens the sense of precise agreement, emphasising each word. Yet ambiguity reasserts itself, as Norrman points out, in passages in which the affirmation confirms a belief in something untrue, whether that belief is real or assumed. For example, Norrman points out an exchange between Charlotte and Prince Amerigo in *The Golden Bowl* in which they discuss the price asked for the eponymous golden bowl by the shopkeeper:

'Is he a rascal?' Charlotte asked. 'His price is so moderate.'  
 She waited but a moment. 'Five pounds. Really so little.'  
 He continued to look at her. 'Five pounds?'  
 'Five pounds.'

However, the reader knows that the shopkeeper actually asked for fifteen pounds. Charlotte's reasons for such a lie, and the Prince's belief or disbelief, remain ambiguous. Responding to criticism that all Jamesian characters sound alike, Norrman suggests that James found it more important to match the form to the context than to individual personalities (although he does not entirely discount the idea that James may have found it difficult to depict different voices).

However, the most fundamental level of James's technique in Norrman's opinion is his use of 'chiastic inversion' which he explains as 'A changes and becomes what B has been; while B changes and becomes what A has been' (Norrman, 1982, p. 138). This structure, he suggests, pervades much of James's writing as well as his thinking. In *The Golden Bowl* he identifies a sequence of such inversions which he illustrates diagrammatically, as shown in Figure 3-2:

**Figure 3-2 Chiastic inversion diagram**



(Norrman, 1982, p. 138)

The diagram illustrates the two couples at the beginning of the story, the Prince and Charlotte (P+C) being most closely linked with Adam and Maggie (A+M) likewise. At Stage 2 the two weddings have occurred and formally the linkages have reversed. However, the four protagonists soon revert to their original

arrangement as Maggie spends most of her time with her father and the Prince and Charlotte develop their affair. Finally, there is a last reversal as Adam takes Charlotte with him to America, and Maggie and Amerigo are left to attempt a real relationship. Chiastic inversion is used at different levels, including humour and individual incidents within the plot as a whole. Norrman brings in a wide range of James's other works, which are beyond the remit of this account. Within *The Golden Bowl*, he finds a mirroring of the chiastic macro-level in antitheses at the micro-level. The characters amuse themselves with this, such as the Prince joking about bringing people out and in (James, 2000 [Original work published 1904], p. 6). The antithesis even extends to the reader's conclusions about the novel. Norrman relates that readers are divided in their loyalties, some favouring Maggie and Adam, others Amerigo and Charlotte. 'The reader's vacillation is endless and the choice impossible' (Norrman, 1982, p. 176). Norrman concludes that insecurity and ambiguity are absolutely fundamental in the Jamesian world, structured as it is in, and by, chiasmus.

Wilson (1981) discusses *The Golden Bowl's* focalization and is sceptical of James's claim, in the Preface to the New York edition of *The Golden Bowl*, that the two books which comprise the novel are narrated successively by the Prince and the Princess. While he concedes that the Prince is the centre of attention of the first book, he points out that others intrude, notably the Assinghams. In fact, James acknowledges this himself; he explains that:

the Prince, in the volume over which he nominally presides, is represented as in comprehensive cognition only of those aspects as to which Mrs Assingham doesn't functionally – perhaps all too officiously, as the reader may sometimes feel it – supersede him (James, 1935, p. 330).<sup>29</sup>

In other words, James accepts that, while the Prince may be the dominant focalizer in the first book of *The Golden Bowl*, there are lengthy passages where we see through Mrs Assingham's eyes. For example, in Chapter 2 there is a description of the Prince looking like 'some very noble personage ... acclaimed by the crowd in the street and with old precious stuffs falling over the sill for his support' (James, 2000 [Original work published 1904], p. 26). This must either be Mrs Assingham's perception if it is not simply the narrator's. The image follows Mrs Assingham's direct speech and a description of the Prince's eyes, also by a classical comparison. The paragraph ends: 'He looked younger than his years; he was beautiful, innocent, vague' (ibid.). All this suggests that the viewpoint is Mrs Assingham's.

In Chapter 3 the Assinghams are alone and the predominant point of view is that of Colonel Assingham. The narrator also intrudes from time to time. For

---

<sup>29</sup> This quote shows that James's late style is also difficult when he is writing non-fiction in the late period, using parenthesis and unusually placed adverbs,

example, in Chapter 25 the narrator comments explicitly on the pagoda metaphor, including using the first person singular pronoun (James, 2000 [Original work published 1904], p. 234). Later in the chapter, as Cohn (2001) points out, the tense used gives us the narrator's knowledge of a later time in the story than Maggie's current position: 'Such things, as I say, were to come back to her' (James, 2000 [Original work published 1904], p. 243). A greater variety of focalization may be useful to rouse and sustain our interest in all the characters, and to maintain the balance between the dilemmas of the protagonists in the symmetrical relationships which Norrman has described. However, the intrusion of the narrator is difficult to explain and runs across James's own accounts of *The Golden Bowl*. Wilson also emphasises the presence of the narrator, who he identifies with James himself. 'The editorial in a James novel is constant in the style and remains, throughout, the reader's vital referent' (Wilson, 1981, p. 66). For Wilson, it is through James that we relate to each character with 'distanced-empathy' (Wilson, 1981, p. 282) which prevents a too-easy moral judgement of the actors. His comparison is with music, with different harmonies working together to form the whole. Contradicting Norrman, Wilson sees the language of the book as speech-like, resembling theatre in its declamatory quality, an effect which is reinforced by James's use of scenes. This produces a hyper-reality 'depicting ultimate human problems more complexly, more variously, more penetratingly and more comprehensively than they had hitherto been depicted' (Wilson, 1981, p. 283).

Bradbury (2000), in her Introduction to the Wordsworth Classic edition of *The Golden Bowl*, takes yet another slant. She particularly focuses on James's 'uncanny and incalculable use of the negative' (Bradbury, 2000, p. viii). She sees this as part of the fundamental structure of the novel – 'a kind of parallel text to the narrative of *The Golden Bowl*, articulating through suppression and hesitation an unspeakable story of betrayal and pain and of cruel repercussions' (Bradbury, 2000, p. viii). She cites examples of Maggie thinking about what hasn't been said, of responsibility for action not acknowledged, and even avoidance by exaggeration. For example, the Prince compares Adam Verver to 'Alexander furnished with the spoils of Darius', exaggerating his father-in-law's wealth and power to turn aside from the humiliation of being collected himself like a work of art. Bradbury is particularly conscious of the effort the reader must make to come to understand a story which is told so obliquely. The task is so difficult, she suggests, that moral judgement is deferred while James's convoluted sentences are negotiated. The result is a meditation on how far one can know another person, and how far one is responsible for their fate, as Charlotte's suffering is mediated through Maggie's focalization and formed by her actions. None of this difficulty is accidental. 'There is a deliberate parallel between "the process and the effect of representation" in the

novel and the drama of the protagonists' (Bradbury, 2000, p. xiv). This is the justification for the late style, for its ambiguity and difficulty.

### 3.4 Conclusion

Any new researcher comes to James's novels conscious of the weight of preceding criticism, which started with James's own revision of his novels. This survey has shown some of the broad range of commentaries, which focus on themes, influences, language and narratorial devices. Views on James's style are quite polarised; I have included an even number of admirers and critics of the late style here, which seems to me to represent the field fairly. There is a striking lack of consensus among critics of James; inevitably a subjective approach to literary analysis results in differing conclusions. My corpus stylistic analysis aims to furnish objectively verified data to be used for Jamesian literary criticism.

Although detailed discussion of James's language exists, it does not predominate. There have been a number of important studies which have included analyses of James's syntax. I have discussed the examination of the syntax of *The Ambassadors* by Watt, Levin and Cross, and Cross's emphasis on James's extreme manipulation of syntactic rules. Leech and Short also discuss syntax, but their study, as with those of Watt and Levin, takes a very small sample of James's texts and any extrapolation to general statements on James's style can only be tenuous. Cross makes general statements in addition to her discussion of the early sentences of *The Ambassadors* but without data to support her conclusions, they are personal impressions, however well informed. A persistent question is to what extent an analysis of a short passage can be taken to exemplify James's style. Yet non-computerised work cannot realistically tackle large amounts of data. It is this problem that this dissertation sets out to tackle. My approach will be explained in the next chapter, which sets out my methodology.

# Chapter 4    Compiling the Corpus: Methodology

---

In order to explore the differences, if any, between Henry James's early and late styles, I have compiled a corpus of parts of one early novel, *Washington Square* and one late novel, *The Golden Bowl*; the reason for this choice of novels is given in section 4.1. I have named this corpus the *Henry James Parsed Corpus (HJPC)*. This chapter explains how I found appropriate e-texts to use for the corpus, which was compiled using a software program called ICECUP, detailed in section 4.2. I will describe the process by which I parsed my corpus along with the extra coding I added to the standard program. Some of the difficulties encountered in the parsing process are set out in section 4.5.

## 4.1    Sourcing the texts

As described in section 3.2.6, Hoover (2007) uses the vocabulary of James's novels, and methodology more often used in author attribution, to identify and characterise James's novels. He finds that they divide quite clearly into three phases of novel writing from 1871-1881, 1886-1890 and 1897-1917. In 1881 James published *The Portrait of a Lady*, and then did not publish another full length novel until *The Bostonians* in 1886. The second gap corresponds with what Edel has called James's 'theatrical years',<sup>30</sup> and follows the commercial failure of *The Tragic Muse*, published in 1890. *The Spoils of Poynton* was published in 1897, beginning the final period of James's writing which culminated in the four major novels of the early 20<sup>th</sup> century.<sup>31</sup> James continued to publish short stories in most years.

As one goal of my project is to identify the differences in syntax between James's early and late styles, I chose a novel from the first and third periods of his publishing career. For the early work, a major concern was finding a text which exemplified the early style rather than a later version revised by James, for example for his New York edition. Between 1907 and 1909 James published this 'monumental edition' in which the novels included were 'revised, updated, rewritten' (Le Fanu, 1982, p. vii), as well as being given lengthy critical prefaces by James himself. Bradbury describes the edition as 'an act of summation and self-presentation designed as both an artistic statement and a commercial enterprise'

---

<sup>30</sup> The title of Leon Edel's doctoral thesis from the University of Paris was 'Henry James: *les années dramatiques*.'

<sup>31</sup> This excludes *The Other House* which was published in 1896 and was an adaptation of a play which had not been produced. Hoover excludes it at least partly because he was unable to find an e-text for it. (Personal communication 19<sup>th</sup> July 2006.)

(Bradbury, 2000, p. VI). Most modern editions of James's novels are based on the New York edition, presumably because it represents the author's definitive last and preferred version of his works.

It was necessary to find an e-text, rather than a printed text, as it would have been impossibly lengthy and complicated to have generated an e-text of my own. Although there are many versions of James's novels freely available on the internet, their quality is very variable. Perhaps the most well-known source of e-texts is Project Gutenberg at [www.gutenberg.org](http://www.gutenberg.org) (Hart, 2010). On this site there is a long list of James's works which are available to download. Unfortunately, there is often no information as to which edition has been used by the volunteer who generated the text. Indeed, I understand that texts are often processed by more than one person, so that they may include material from several editions. Clearly, this site was not a suitable source for my data. There are a large number of other internet sources for Jamesian e-texts, ranging from the academic to the popular, but all require caution when the date of the source text is an important issue. Detailed referencing is an exception, rather than the usual practice. It is possible to view electronic facsimiles of the original publication of *Washington Square* in Cornell University's facsimile of Harper's *New Monthly Magazine*, Volume LXI (Cornell University Library, 2010), but this is not computer-readable for the purpose of compiling a database. The vast 'Henry James scholar's guide to websites' (Hathaway, 2010) lists a wide range of sources for electronic texts, as well as other useful links. It includes a number of novels (and other works) in electronic form prepared by Richard Hathaway, which have careful references to the exact editions used. Often, however, this is the New York edition, or an earlier but revised version, rather than the originally published text. The version of *Washington Square* on Hathaway's site is from 'the London edition (Macmillan, 1881), for which James made his final revisions' (Hathaway, 2010). It is this version which I have used for my first corpus text. An earlier novel might have been a preferable choice, but all the other e-texts available were based on the New York edition. However, *Washington Square* was excluded by James from the New York edition, which means that the latest text of the novel available dates from the end of his first period of novel writing.

For the late novel, it seemed reasonable to choose James's last complete work, *The Golden Bowl*. In this case, the e-text I am using is from the New York edition, but this can only enhance the 'lateness' of the late style used. The web addresses of my two texts, for *Washington Square* and *The Golden Bowl* respectively, are <http://www2.newpaltz.edu/~hathawar/washsq.html> and <http://www2.newpaltz.edu/~hathawar/goldenbowl1.html>.

## 4.2 ICECUP

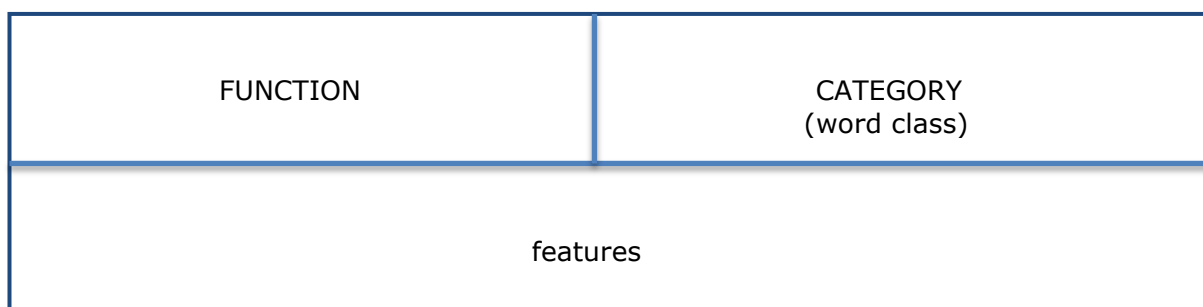
I compiled the *HJPC* using the ICECUP 3.1 program (Nelson, et al., 2002a). This corpus exploration program was developed at the Survey of English Usage at University College London as part of the International Corpus of English (ICE) project. ICECUP (the International Corpus of English Corpus Utility Program) was developed by Sean Wallis in order to process and display the data contained in ICE-GB, which is the British component of the wider ICE project. ICE was started by Sidney Greenbaum in 1988, when he was Director of the Survey, and aims to bring together corpora of a number of different varieties of English from around the world. ICE-GB has been a valuable reference source while parsing Henry James's sentences. As far as possible I have followed the grammar used in ICE-GB, which is broadly that of Quirk et al. (1985) for the *HJPC*. My requirement was simply for a reliable framework which I could apply consistently across the data in order to make comparisons between different texts. Quirk et al's grammar provided such a framework as well as having the advantage that ICECUP was designed to work with it, and that ICE-GB provided a corpus parsed by the same criteria which I could refer to resolve difficult questions of parsing in a consistent way. The grammar used for ICE-GB has dictated many aspects of the analysis in my corpus, such as my handling of coordinated clauses and phrases, categories of dependent clauses and types of transitivity. Although the syntactic analysis in the grammar may be debatable in some areas, for example the exclusion of complements from the verb phrase, such problems have not concerned me.

ICECUP displays the sentences of the corpus in a 'tree' format with a great deal of syntactic information about each sentence. Nelson et al. explain (2002b, p. 22):

Each node on the tree consists of the three sectors [...]. The function and the category/word class sectors are always labelled, but on many nodes the features sector is blank. In many cases no features are applicable. Function, category, and word class labels are shown in upper case. Feature labels always appear in lower case.

This node structure is shown in Figure 4-1:

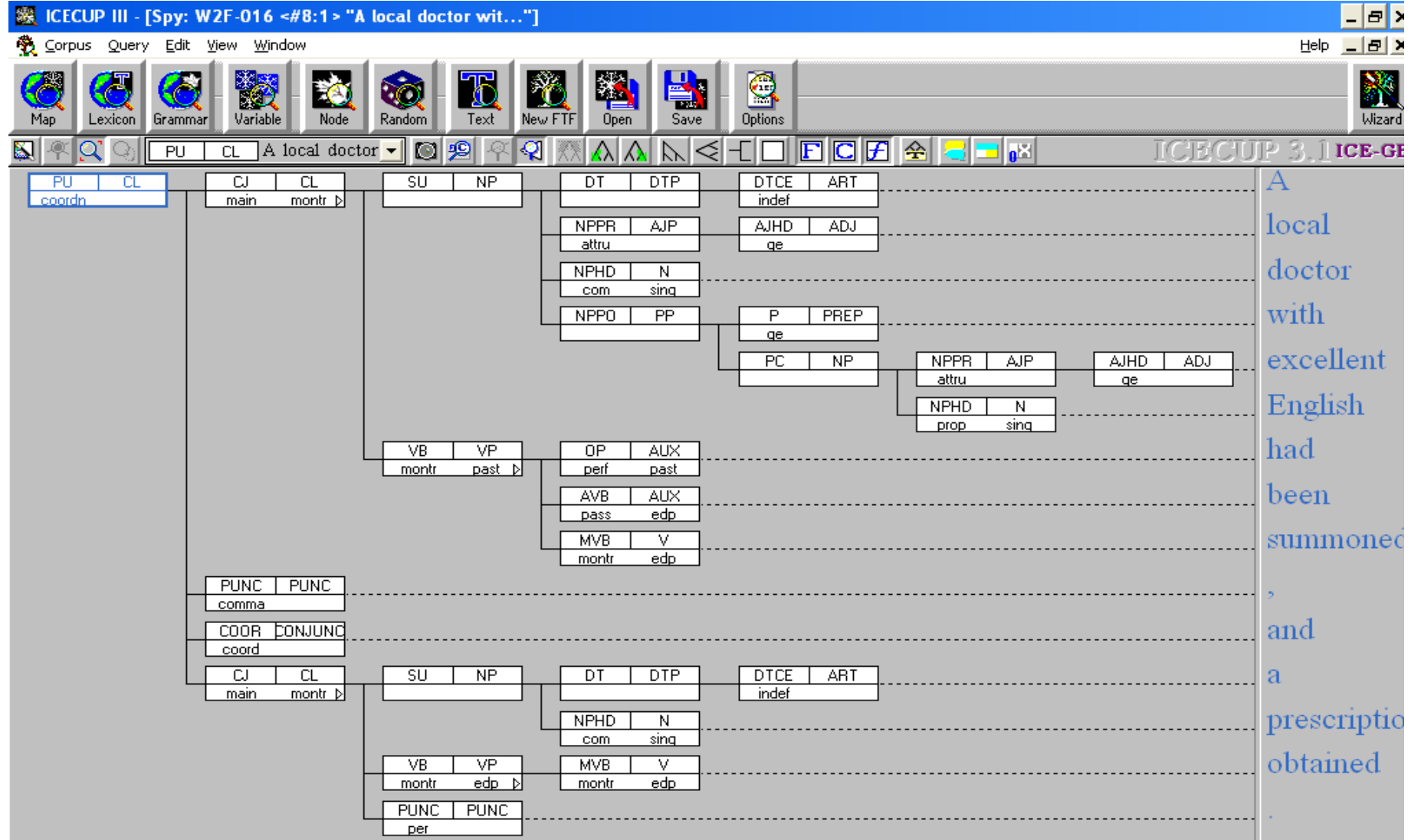
**Figure 4-1 The sectors of a node (Nelson, et al., 2002b, p. 22)**





An example of an ICE-GB sentence can be seen in Figure 4-2. The sentence is shown at the bottom of the page and also down the right margin, so that each word has a separate line of analysis. The tree starts at the top left hand corner where the first node indicates that this is a Clause (CL) functioning as a Parsing Unit (PU). Every node in ICECUP indicates in this way the function of the item on the left and its form on the right, as explained above. In the lower portion of the node, in lower case letters, are further features of the item. In this case the feature indicates that this clause is coordinated. The CL-node stands alone vertically, dominating the entire sentence. In the next slot to the right, reading vertically, there are four nodes. These show the two halves of this sentence, which consists of two conjoined clauses (CJ/CL), a coordinating conjunction (*and*) and an item of punctuation – a comma. Continuing to the next column to the right, the nodes from top to bottom show the elements of each clause – each has a subject and verb (here shown as VB – verbal) and the final full stop is shown at the bottom. The subject of the first clause is broken down into a determiner phrase, an adjective phrase which premodifies the noun *doctor*, a noun and a prepositional phrase which postmodifies the noun. (A list of the abbreviations used in this sentence is included after the diagram, and there is a complete list of all the ICECUP abbreviations used in this dissertation on page 13.) For each of these elements there is at least a two level analysis; for example the word *local* is shown to have the function of a Noun Phrase Premodifier (NPPR) and the form of an adjective phrase (AJP). It is then defined, in the most right hand box as functioning as the head of the adjective phrase with the form adjective. The verb phrase (VP) in the first clause has three elements: it consists of two auxiliary verbs, *had* and *been*, which are assigned the functions of operator (OP) and auxiliary verb (AVB), and a main verb (MVB) *summoned*. The features in smaller print beneath each verb give information as to the tense, passive mood and transitivity of these verbs which then percolate to the clause level as can be seen in the Conjoin/Clause (CJ/CL) node at the top of the page. The small triangles in some of the feature nodes indicate that there is more information which will only be shown if the node is expanded further on a computer screen.

Figure 4-24-2 An ICE-GB sentence displayed in an ICECUP tree



A local doctor with excellent English had been summoned, and a prescription obtained.

## Abbreviations used in Figure 4-2 (in alphabetical order)

ADJ	adjective
AJHD	adjective phrase head
AJP	adjective phrase
ART	article
attru	unmarked attributive
AUX	auxiliary
AVB	auxiliary verb
CJ	conjoin
CL	clause
com	common
CONJUNC	conjunction
COOR	coordinator
coord	coordinating
coordn	coordination
DT	determiner
DTCE	central determiner
DTP	determiner phrase
edp	-ed participle
ge	general
indef	indefinite
main	main clause
montr	monotransitive
MVB	main verb
N	noun
NP	noun phrase
NPHD	noun phrase head
NPPO	noun phrase
postmodifier	
NPPR	noun phrase
premodifier	
OP	operator
P	prepositional
PC	prepositional
complement	
per	period
PP	prepositional phrase
PREP	preposition
prop	proper
PU	parsing unit
PUNC	punctuation
sing	singular
SU	subject
V	verb
VB	verbal
VP	verb phrase

### 4.3 The Henry James Parsed Corpus – ICECUP adapted

The *HJPC* consists of 2 parts. One part (*WS*) was compiled from chapters 1, 2, 16, 18 and 31 of *Washington Square*. The other part (*GB*) uses data from chapters 1, 2, 20, 22 and 38 of *The Golden Bowl*. The abbreviations *WS* and *GB* refer to each part of the *HJPC* corpus, while the actual novels are referred to by their full names. There are 8,888 words from *Washington Square* and 23,999 from *The Golden Bowl*, making 32,887 words in all. These five chapters (the first and second chapters, the central chapter and the chapter which falls two chapters before the centre, and the chapter five chapters from the last) were chosen to provide a spread of data across each novel. I wished to avoid only including any particular style which might typify the beginning, middle or end of this (or perhaps any) novel. I would have liked to analyse more data, but the onerousness of the data preparation precluded that. The chapters were chosen before any examination at all of their content or structure, and so the choice was not prejudiced by any considerations of that kind. Other chapters could have been chosen and beyond the wish for a widely-spread sample the choice of chapters is quite arbitrary, but means that, despite the different number of chapters in each book (35 in *Washington Square* and 42 in *The Golden Bowl*), the *HJPC* includes samples of the early development, central developments and final resolution of each plot.

Having chosen the texts and downloaded the e-texts, as described in section 4.1, they had to be prepared by passing them through Maketag (Wallis, 2008), a computer program which adds part of speech ‘tags’ to each word. This is a multistep procedure: a chapter of the e-text must be pre-processed by the program and then corrected before the tags can be attached. The first stage divides the text into numbered sentences but assumes a sentence ends with every full stop, so that *Mr.<sup>32</sup>,* for example, would trigger a new sentence. Once the sentences are correctly identified and numbered and any extraneous material removed (some texts include page numbers from the original publication), then the program can be run to assign tags. Checking that sentence boundaries are in the correct place required a decision as to what a ‘sentence’ should consist of. A sentence here is a piece of text which begins with a capital letter and ends with a full stop. As will be seen below, this sometimes includes more than one clause which would be grammatically acceptable as a free-standing sentence, but I have chosen to take James’s decision as to what should be included in one sentence as part of his style. Once Maketag has done its

---

<sup>32</sup> James adheres to the American usage, putting a full stop after Mr, Mrs and Dr.

work, the program MAKEICE<sup>33</sup> transforms the data into the tree presentation, which can then be edited in ICECUP.

The tagging program attempts to assign parts of speech and function labels at the word level, but at this stage trees are presented completely flat, as in Figure 4-4. Also many of the nodes are incomplete and some may be incorrect. Hand correction node by node was necessary. Some clause structure was added, but the phrase structure was only included if it was likely to be of particular significance, or to clarify a structure. For example, in a sentence such as:

2) A dull, plain girl she was called by rigorous critics (Part of WS03:7)<sup>34</sup>

*a dull, plain girl* is the object complement of *called* which has been moved to the beginning of the clause (or preposed). This is marked in the features of the clause by the abbreviation 'preco' and the phrase is also given a phrase level node to show it is the object complement despite its unusual position, as shown in Figure 4-3.

**Figure 4-3 Part of WS03:7 illustrating phrase node use**

CO	NP	DTCE	ART	A
		indef		
		AJHD	ADJ	dull
		ge		
		PUNC	PUNC	,
		comma		
		AJHD	ADJ	plain
		ge		
		NPHD	N	girl
		com	sing	
				she
NPHD	PRON			
pers	sing			
				was
OP	AUX			
pass	past			
				called
MVB	V			
cxtr	edp			
				by
P	PREP			
ge				
				rigorous
AJHD	ADJ			
ge				
				critics
				—
NPHD	N			
com	plu			
PUNC	PUNC			
dash				

<sup>33</sup> MAKEICE is the indexing program for ICECUP.

<sup>34</sup> Sentences are identified by their part of the *HJPC*, chapter and sentence number. Thus WS03:7 is the seventh sentence in the third chapter of *Washington Square*.

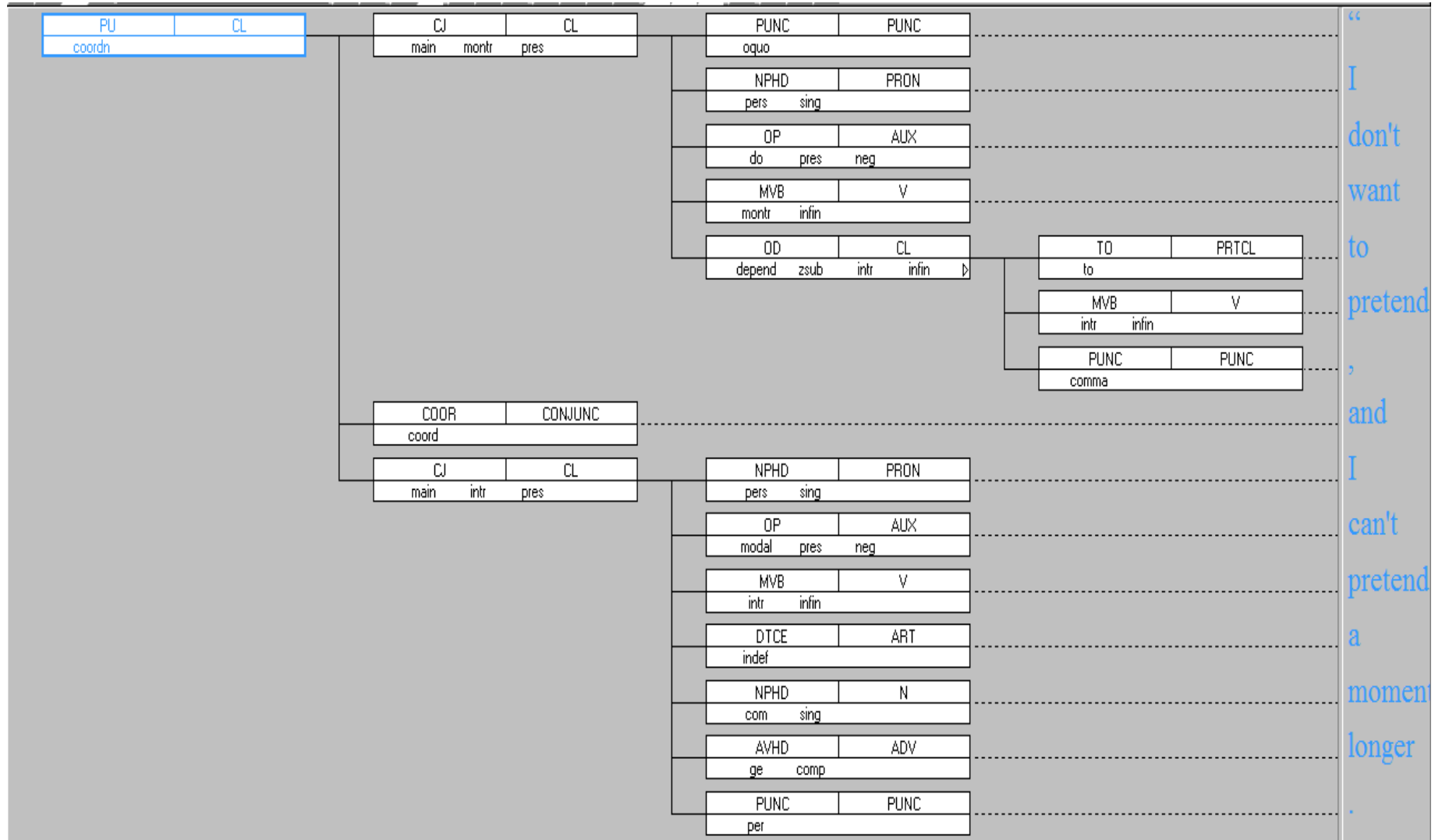
GB05:3 is shown in Figure 4-4 with the structure it had at the end of processing with Maketag and MAKEICE. The final version of GB05:3 after hand processing can be seen in Figure 4-5. Comparison of Figure 4-5 and the ICE-GB sentence shown in Figure 4-2 shows that GB05:3, a sentence from the *HJPC*, has a much simpler tree diagram, although with the same basic conjoined sentence structure.

Figure 4-4 GB05:3 before hand processing

0	0	PUNC	PUNC	“
		oquo		
	0		PRON	I
	pers	sing		
	0		AUX	don't
	do	pres neg		
	0		V	want
	montr	infin		
	0		PRTCL	to
		MVB	V	pretend
	montr	infin		
		PUNC	PUNC	,
		comma		
	0		CONJUNC	and
	coord			
	0		PRON	I
	pers	sing		
	0		AUX	can't
	modal	pres neg		
		MVB	V	pretend
	montr	infin		
	0		ART	a
	indef			
		NPHD	N	moment
	com	sing		
		<b>Features:</b> common; singular	ADV	longer
	ge	comp		
		PUNC	PUNC	.
		per		

“I don't want to pretend, and I can't pretend a moment longer.

Figure 4-5 GB05:3 after hand processing



"I don't want to pretend, and I can't pretend a moment longer."



## 4.4 Specialised notation

In addition to using the standard tagging and parsing categories available in ICECUP and used in ICE-GB, some extra notation has been added to the *HJPC* to mark features of James's text which could then be easily located and quantified. This extra notation, which is called *markup*, is a code additional to the standard ICECUP format and is inserted into the sentence. It can be searched for using the ICECUP program but does not show in the tree diagram. One example is the occurrence of direct speech, whose definition and delineation is described in section 4.4.1.

My research question centres on the analysis of degrees of 'difficulty' in James's novels. However, 'difficulty' is a subjective description, which describes a personal response when reading a text. For an objective analysis, I have sought to characterise and quantify syntactic complexity in the *HJPC*. This issue is discussed in more detail in section 5.2, and the components of complexity at the sentence level are enumerated in section 6.4. In addition, two noticeable characteristics of James's prose were identified during the parsing process<sup>35</sup> and marked for analysis; they are *additional parsing units (APU)* and *delay*. They are defined and described in sections 4.4.2 and 4.4.3 respectively.

### 4.4.1 Direct speech

When looking at the complexity of James's late style, the prevalence of direct speech was likely to be an important factor. Direct speech is different in structure, and may be thought by authors to be simpler in structure, than narrative (including the various types of reported speech). Greenbaum and Nelson (1995) discuss the common belief that speech is simpler than writing in syntactic structure in the preamble to their comparison of the syntax of speech and writing (see section 5.2.3). Therefore, in measuring the markers of complexity in a text, it is important to know how much of the text is direct speech. For this reason, I marked all sentences containing direct speech, that is sentences containing speech marks, with the markup code 'ds'. The majority of these sentences also contain non-speech material, but I have not differentiated these or quantified the proportion of speech to non-speech within an individual sentence. Examples are:

- 3) "Tell him you are to be married." (WS31:24)
- 4) "She says I have a genius for consolation," said Mrs. Penniman. (WS31:9)

---

<sup>35</sup> Like Semino and Short (2004) (see section 2.8) I found that the process of parsing was helpful for identifying characteristics of the text which warranted further investigation.

- 5) "So I am," said Catherine, softly; and she closed the door upon her aunt. (WS31:25)

Speech marks occasionally appear in passages of indirect thought; I have also marked these as direct speech. For example:

- 6) "Why handsome?" Maggie would have been free to ask; since if she had been veracious the service assuredly wouldn't have been huge. (GB38:21)

Very occasionally, James uses a speech-like structure in reported thought but does not use speech marks. I have not marked these as direct speech.

- 7) What retarded evolution, she asked herself in these hours, mightn't poor Charlotte all unwittingly have precipitated? (GB38:42)

#### 4.4.2 Additional parsing units (APU)

As mentioned in section 4.3, James quite frequently writes sentences which contain more than one clause which could stand alone syntactically as a sentence, joining those sentences with different kinds of punctuation, but not with a coordinator as in a conventional compound sentence. I have treated these as one sentence, as it is clearly a feature of James's style to bring linked ideas together in this way. As I wanted to be able to quantify James's use of this construction, I have marked each additional potentially-independent clause with the markup code <apu> (additional parsing unit). Examples (with the APU markup shown in grey) are:

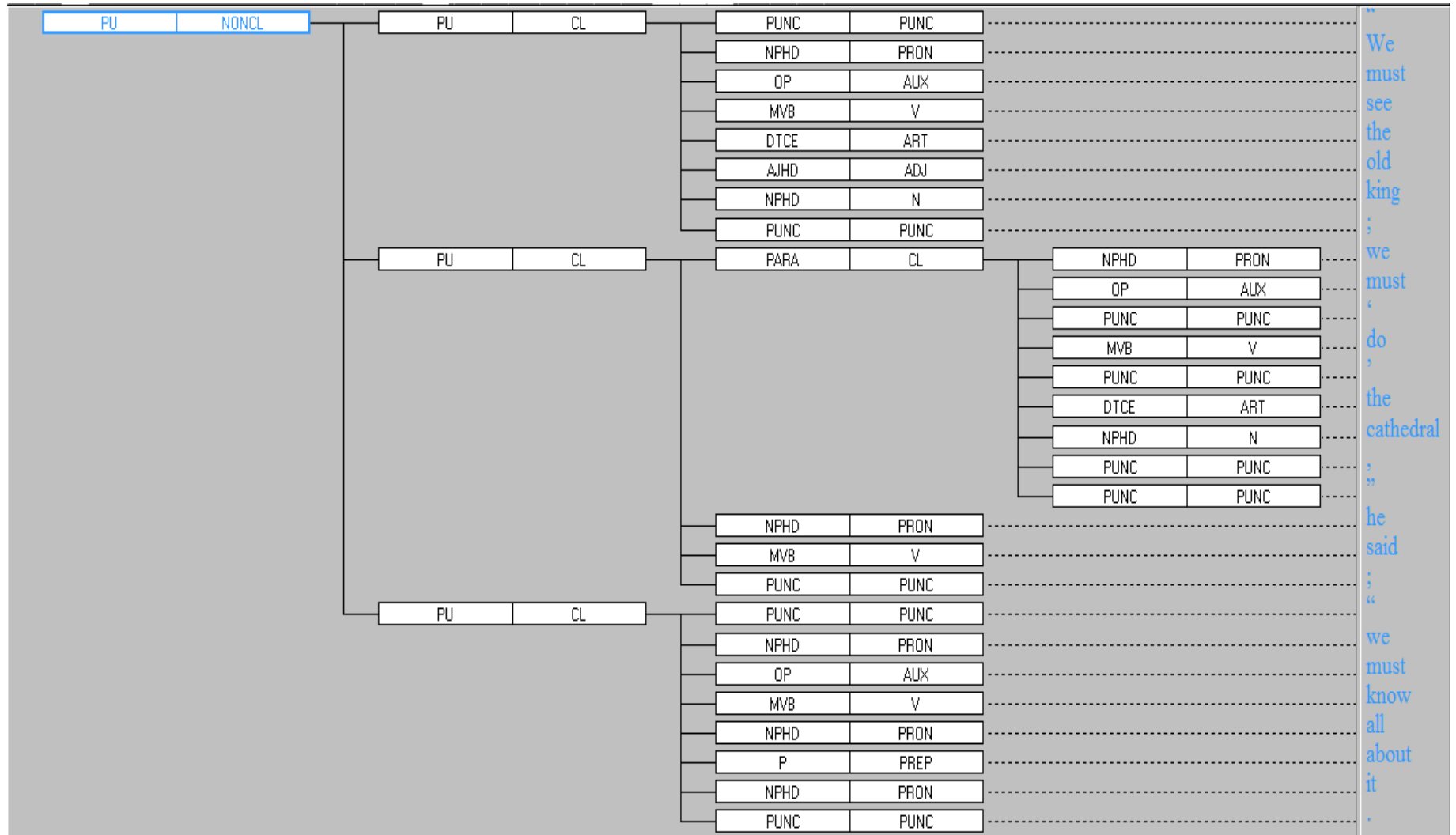
- 8) These at any rate for the present were unanswerable questions; <apu> all that WAS before him was that he was invested with attributes. (GB01:276)
- 9) He had paid first and last many an English country visit; <apu> he had learned even from of old to do the English things and to do them all sufficiently in the English way; <apu> if he didn't always enjoy them madly he enjoyed them at any rate as much, to all appearance, as the good people who had in the night of time unanimously invented them and who still, in the prolonged afternoon of their good faith, unanimously, even if a trifle automatically, practised them; yet with it all he had never so much as during such sojourns the trick of a certain detached, the amusement of a certain inward critical, life; the determined need, while apparently all participant, of returning upon himself, of backing noiselessly in, far in again, and rejoining there, as it were, that part of his mind that was not engaged at the front. (GB20:10)
- 10) "You ask if I'm likely also to back out then, because it may make a difference in what you and the Colonel decide?"-- <apu> he had gone as far as that for her, fairly inviting her to assent, though not having had his impression, from any indication offered him by Charlotte, that the Assinghams were really in question for the large Matcham party. (GB20:20)

Note that in GB20:10, shown as example 9), the third semi-colon after *them* is followed by the coordinator *yet*, so I have not treated the following clause as an

APU but as a coordinated main clause. (This sentence is also a good example of the long and complex sentences of the late style.) The fourth semicolon, after *life*, has also not triggered a new main clause, but rather a second part of the direct object of *had* of which the first is *the trick*. While the first two sentences have semi-colons as markers of the new main clause, in GB20:20 a dash is used instead; there seems to be no pattern in the punctuation choices, though that might be an area for further research.

The use of APU is illustrated in Figure 4-6, which shows GB22:67. Note that this tree has been contracted so that the features below the main labels are not shown, so as to fit it onto the page. Although the tree diagram is similar to those seen before, the sentence text at the bottom of the page shows additional code, and the coded features are also echoed in some features of the tree diagram. For example, the overarching Parsing Unit, shown in the top left corner, is not a clause but a non-clause (NONCL). This is because this Jamesian sentence contains three canonical sentences, which are each given their own parsing unit, shown in the column immediately to the right of the overarching node. These are also marked in the sentence text with the code <apu>. As this sentence contains direct speech, the code <ds> is at the beginning of the sentence. The codes <0!> and </0!> simply indicate that the parsing of this sentence is complete.

Figure 4-6 GB22:67 showing markup



<#67:1> <sent> <ds> <O!> </O!> “We must see the old king; <apu> we must ‘do’ the cathedral,” he said; <apu> “we must know all about it.

### 4.4.3 Delay

Many of the critics discussed in Chapter 3 mention James's frequent use of parenthesis. It is a striking feature of his style. There is a bracketed parenthesis in the sixth sentence of *Washington Square*:

- 11) Though he was felt to be extremely thorough, he was not uncomfortably theoretic; and if he sometimes explained matters rather more minutely than might seem of use to the patient, he never went so far (like some practitioners one has heard of) as to trust to the explanation alone, but always left behind him an inscrutable prescription. (WS01:06)

and a parenthesis marked with dashes in the seventh sentence of *The Golden Bowl*.

- 12) He was too restless -- that was the fact -- for any concentration, and the last idea that would just now have occurred to him in any connexion was the idea of pursuit. (GB01:07)

A reader of either novel would notice this characteristic, and might find it adds to the difficulty of reading James, especially if the parenthesis is much extended. There is a certain feat of memory required, to a greater or lesser extent, for the reader to remember the first part of the sentence when the second appears after the parenthetical interruption. My thesis is that this may be part of the complexity of James's late style, and that there may be a difference in the frequency and/or amount of parenthesis between *Washington Square* and *The Golden Bowl*. For this reason I have devised a method to quantify delay, which I define in this way:

Any element inserted between the nodes of a kernel clause, defined as Subject + Predicator + (Indirect object) + (Direct object) + (Object complement), occasions delay, which is quantified by the number of intervening words.

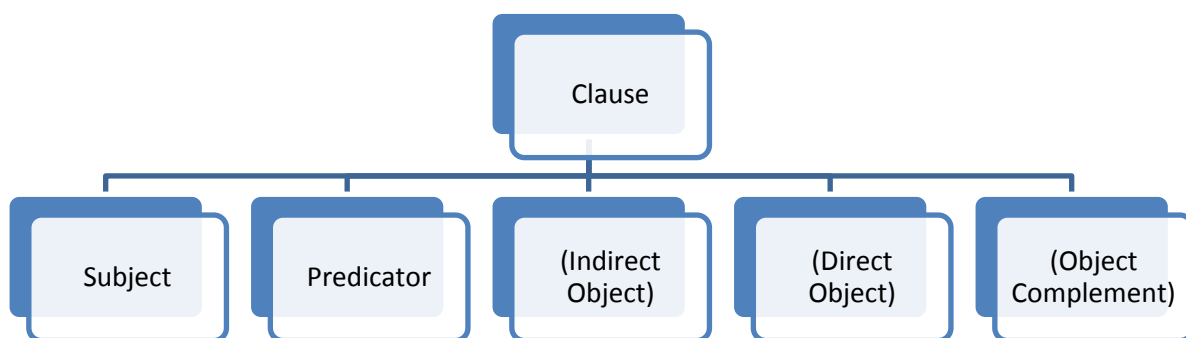
The elements in brackets might, or might not, be present in a kernel clause depending on the transitivity of the predicator.

To expand on this definition, the notion of a kernel clause is adapted from Huddleston (Huddleston, 1984, pp. 13-16), but Huddleston uses the term to refer only to main, declarative, positive and non-reduced clauses, whereas I include subordinate, interrogative and negative clauses as well as those including ellipsis. Whereas Huddleston is interested in using the term 'kernel clause' to characterise the clause as a whole as compared to possible other types of clauses, I am interested in the internal structure of the clause. A canonical or kernel main clause can be defined as consisting of:

Subject - Predicator - (Indirect Object) - (Direct Object) - (Object Complement)

where the elements in brackets may or may not be present. (Huddleston, 1984, pp. 177-8). This definition also derives partly from the discussion in Aarts (Aarts, 2008, pp. 8-9). The term 'Object Complement' comes from the ICECUP grammar (Nelson, et al., 2002b, p. 51). The structure of a kernel main clause is illustrated graphically in Figure 4-7:

**Figure 4-7 Structure of a kernel main clause**

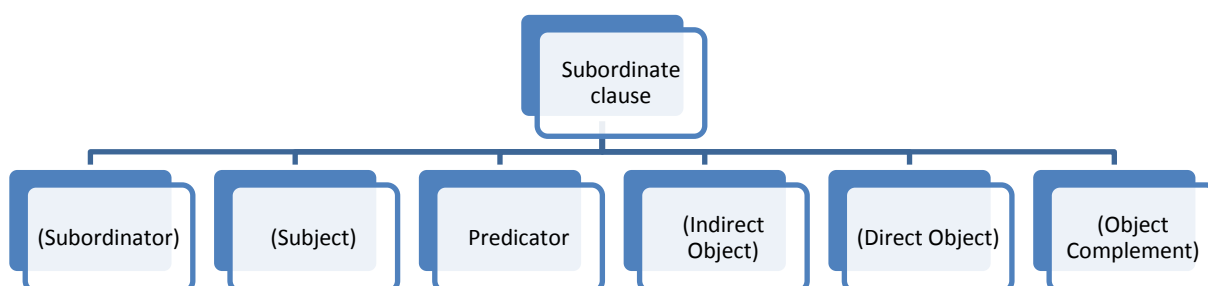


A canonical or kernel subordinate clause can be defined as consisting of:

(Subordinator)-(Subject)-Predicator-(Indirect Object)-(Direct Object)-(Object Complement)

This is shown in Figure 4-8:

**Figure 4-8 Structure of a kernel subordinate clause**



It is a feature of the ICECUP grammar that it treats the verb phrase as separate from its complements, in line with Quirk *et al.*:

Verb phrases consist of a main verb which either stands alone as the entire verb phrase or is preceded by up to four verbs in an auxiliary function. (Quirk, et al., 1985, p. 62)

This is in contrast to Huddleston and Pullum (2002, p. 22), who state that 'a verb and its various complements make up a verb phrase', although in his earlier *Introduction to the Grammar of English*, Huddleston used the term predicate to mean a predicator plus its complement(s), but also used the term *extended verb phrase* and *verb phrase* with the same meaning. Here I have used 'predicator' to refer to the function of the verb, and have assumed that all complements are separate both in form and function.

In both main and subordinate kernel clauses, adverbial clauses or adverb phrases can be placed between these constituents, but then the reader has to keep in mind the syntactic link so as to parse the sentence and understand the relationships it represents. For example, in example 13) the markup code in grey shows that the subject complement (CS) is delayed by 4 words. The delay is occasioned by the insertion of the phrase *for about five years* between the verb *were* and its complement *a source of extreme satisfaction*. The sentence is grammatically correct but includes delay which could have been avoided with no change of meaning by placing the phrase *for about five years* at the beginning of the sentence. Similarly, in example 14) the direct object (OD) *a certain sense of duty* is delayed by the five words of *in his own narrow way* which could have been placed at the end of the sentence.

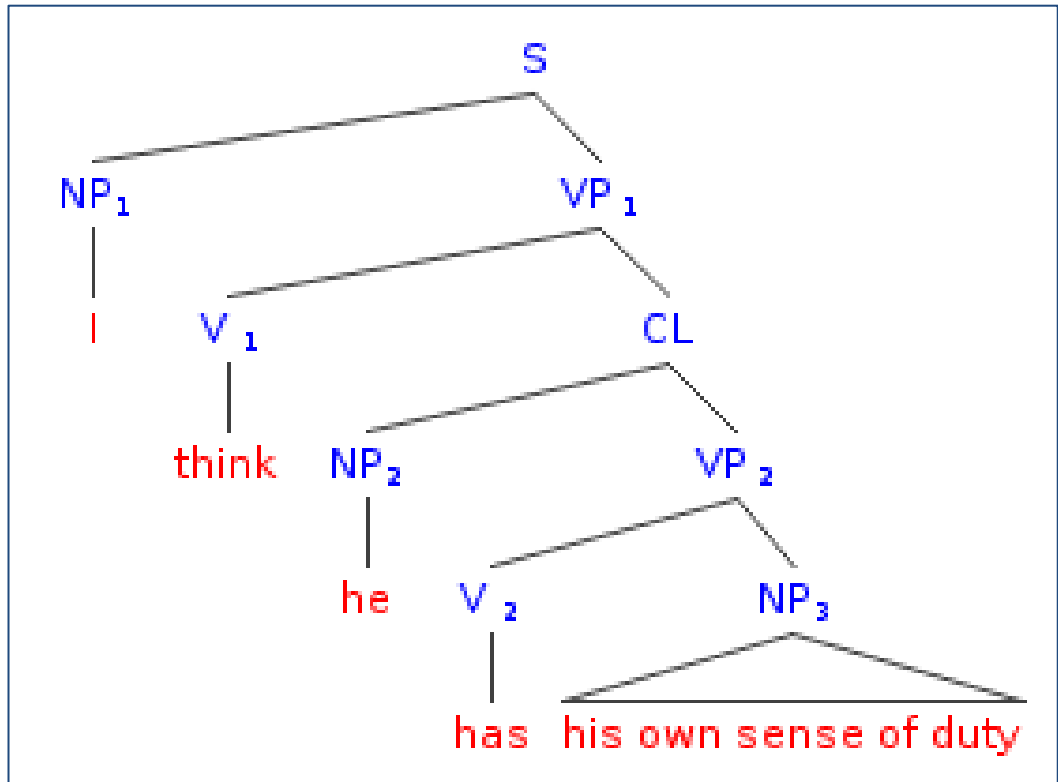
- 13) These eyes, and some of their accompaniments, *were* for about five years <delay> <4> <CS> *a source* of extreme satisfaction to the young physician, who was both a devoted and a very happy husband. (WS01:18)
- 14) "He is not sentimental," said Mrs. Penniman; "but, to be perfectly fair to him, I think he *has*, in his own narrow way, <delay> <5> <OD> *a certain sense of duty*." (WS16:90)

Sentences like examples 13) and 14), where the subject complement and direct object, respectively, are delayed, present particular theoretical challenges in a system where the verb phrase is considered to include both the verb and its complements. In a phrase structure grammar which holds that only binary branching is possible, as shown in Figure 4-9, there is no place to insert the parenthetical prepositional phrase *in his own narrow way* into the second VP (VP2) *has his own sense of duty*. Theorising which grammar better represents language processing is beyond the scope of this project. However, if Figure 4-9 represents the way sentences are processed by a reader, it may be that such delays are particularly likely to add to a perception of a text being difficult.<sup>36</sup> In section 5.2.4 I discuss the impact of a small number of exceptional sentences on perceptions of reader difficulty. This may be linked to the disruption of syntactic norms in many of those sentences.

---

<sup>36</sup> I am indebted to Dr Joanne Close for her help with this idea. The phrase structure tree in Figure 4-9 was drawn using phpSyntaxTree website (<http://ironcreek.net/phpsyntaxtree/>).

Figure 4-9 Phrase structure tree



James's style is characterised by the positioning of adverbial clauses or phrases in unusual places, in addition to the parenthetical interruptions illustrated in examples 11) and 12). Thus, my definition of delay is both broader than, and includes, parenthesis. As adverbials can be placed in a wide variety of positions, I do not consider them to be capable of being delayed. For example:

- 15) He desired experience, and in the course of twenty years he got a great deal. (WS01:24)

In this sentence, if the adverbial prepositional phrase *in the course of twenty years* was moved to the end of the sentence, it would not be delayed; it would merely have been placed in an alternative position. Figure 4-10 illustrates how delay has been marked and quantified in GB22:65, which is given here as example 16):

- 16) And *her eyes, which*, though her lips smiled, *were* almost grave with their depths of acceptance, *came* back to him. (GB22:65)

Delay is only shown in the tree diagram by the insertion of phrase level nodes for the subject *her eyes*, and its corresponding predicator (in ICECUP terms: VB/VP for Verbal/Verb Phrase) *came* in the main clause, and the subject *which* and its predicator *were* in one of the subordinate clauses. In the sentence text at the bottom of Figure 4-10, the markup code is shown both detailing and quantifying the delay. For example, after *smiled* is the code <delay> <4> <VB>. This indicates



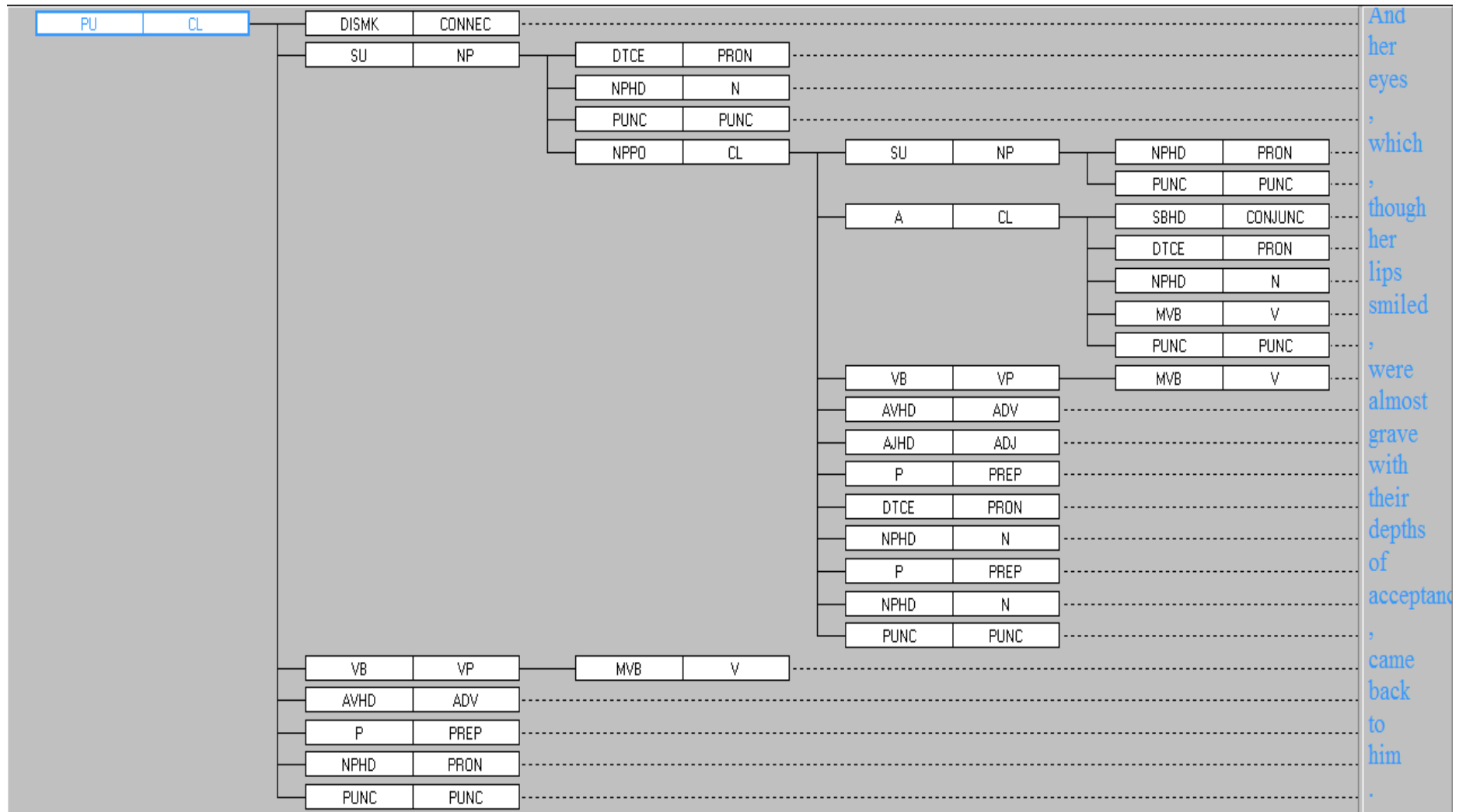
that there has been a delay at this point, that there are 4 words between one element of the canonical sentence (here the subject) and the next syntactically linked element, and that the item which has been delayed is the predicator, shown by the ICECUP code VB.

The relative clause *which were almost grave with their depths of acceptance* falls between the subject of the main clause *her eyes* and the main verb *came*. This presents a new problem. Although this sentence is very typically Jamesian in its parenthetical interruption between *eyes* and *came back to him*, syntactically the relative clause is part of the subject within which the noun *eyes* is postmodified by a non-restrictive relative clause. This means that there is no delay according to my preceding definition between the subject noun phrase and the predicator. However, as the effect of this construction is clearly very similar to that of formal delay, I have included it under that heading and modified my rule as follows:

Any element inserted between the nodes of a kernel clause, defined as Subject + Predicator + (Indirect object) + (Direct object) + (Object complement), occasions delay, which is quantified by the number of intervening words. Delay is also occasioned when an element is postmodified by a clause.

To return to Figure 4-10, there are 13 words between the subject *eyes* and the predicator *came* and this is noted in the sentence as before. For convenience, the total *delay score* is indicated at the beginning of the sentence for easy data collection. This is simply the total of all the separate instances of delay in the sentence, which here is 17 (shown as D17). When counting words for Delay, my definition of a word follows that of ICECUP. Usually words are defined orthographically but some are linked (the ICECUP term is ditto-tagged) and counted as one word. This category includes proper names (*Mrs. Verver, Charlotte Stant*), compound prepositions (*in addition to*), connective expressions (*above all*) and nouns modifying nouns (*charity performance*). It includes some words which would nowadays be one word, although James writes them separately, such as *none the less*. [All examples are from the *HJPC*.]

Figure 4-10 GB22:65 showing delay



<#65:1> <sent> <D17> <O!> </O!> And her eyes, which, though her lips smiled, <delay> <4> <VB> were almost grave with their depths of acceptance, <delay> <13> <VB> came back to him.

ICECUP has a category which overlaps partly with delay. Some clauses and phrases are marked as *detached functions*. For example:

17) ‘You’d better not, Brett. (W2F 001:63)<sup>37</sup>

18) P. W. Botha, the South African premier, was to resign. (W2F 003:12)

The tree diagram of example 17) shows the noun phrase *Brett* as a detached function with the feature *vocative*, as it is separate from the syntax of the sentence, whereas in example 18) the phrase *the South African premier* is a detached function used appositively. Whereas 17) does not represent delay, in 18), because the first in a pair of appositive phrases is always considered in ICE-GB to be the syntactic node and the second the detached function, delay would be counted between the subject (*P. W. Botha*) and the predicator. The detached function category also applies, as in Figure 4-11, to comment clauses inserted into a sentence of direct speech, such as:

19) He *made*”, she said, “*a great impression* on me”. (GB22:79)

This illustrates delay between the predicator and its direct object, caused by the comment clause. However, it is more usual for the clause to be placed so that it does not occasion delay, as in example 20):

20) “Then you had thought,” he wondered, “about Gloucester?” (GB22:107)

In this sentence the prepositional phrase *about Gloucester* is an adverbial placed at the end of the canonical sentence, not interrupting it. As explained using example 15), adverbials, including adverbial prepositional phrases, can be placed in a number of positions in a sentence, so no delay is counted. Although there is still some effort of memory required on the reader’s part, it is vital for the greatest possible consistency of analysis that there is a theoretical basis for labelling and quantifying features. Inevitably some ambiguous examples exist, examples of which will be discussed in Section 4.5. Some detached functions realised by comment clauses would fit Mahlberg’s definition of suspensions – see Section 2.9.2 – if the comment clause is a minimum of 5 words. However, the function of these shorter insertions into the James’s sentences seems rather to be to foreground the words which follow the interruption rather than to characterise the speaker as Mahlberg found in Dickens’s works.

---

<sup>37</sup> ICE-GB sentences are identified by a code which indicates their text type; *W2F* indicates that examples 17) and 18) are from the written part of the corpus, within the creative writing section and are taken specifically from novels/stories. *001* and *003* indicate the text number and the figures after the colon show the sentence number.

Espinal (1991) analyses similar phenomena, which she calls 'disjunct constituents'. She includes in this group a wide variety of syntactic categories from noun phrases and adverbial clauses to whole interpolated sentences. They are identified as disjunct because they 'bear no obvious syntactic relationship to the sentences they seem to be included in' (Espinal, 1991, p. 726). Espinal takes a Generative Grammar approach and attempts to formulate a multi-language theory which accounts for this phenomenon. She includes in her data three types of disjunct: those which are linked to the main clause pronominally, those 'that contain a syntactic gap filled conceptually by the main clause' (Espinal, 1991, p. 729), such as *I think*, and those which are syntactically completely independent of the main clause. All three types would be likely to constitute delay in my system, depending on their placement within the sentence. Espinal suggests that a three-dimensional structure is necessary to integrate disjuncts into syntactic theory; the details of her solution are beyond the scope of this dissertation. She does not comment on any difficulty these 'simultaneous phrase structures [which] codify autonomous syntactic information' (Espinal, 1991, p. 759) may cause the reader or listener.

Blakemore's paper (2009) on free indirect style is an example of a different theoretical approach to delay, discussing and illustrating how classifications of this phenomenon differ. Blakemore contrasts different linguistic descriptions of parenthesis, which may be seen as a pragmatic phenomenon, occasioned by hesitation or self-correction, or a syntactic one, which would include such features as non-restrictive relative clauses and apposition. Blakemore's approach is pragmatic and relates to the way the narrator is revealed by parentheticals within free indirect thought and speech. She also contrasts fluency and disfluency, both of which may or may not be syntactically licensed. Fluency involves a comment on something within the sentence, whereas disfluency represents a digression in thought.

My definition of delay includes both syntactically licensed and unlicensed parentheses, which may be fluent or disfluent. I have defined it purely syntactically but some examples still require a subjective judgment. This is because some words or phrases can only be placed between the nodes of a canonical sentence. A particularly clear case is the adverb *never* in a sentence like *I have never written a book*. Although *never* is placed between the auxiliary and main verb, it cannot be counted as delay as there is no other syntactically licensed place for it to occur in the sentence.

- 21) When you **don't**, at least," he had amended with a further thought, <delay>  
<8> <MVB> "see too little." (GB01:117)

- 22) "I don't *feel*, my dear, if you really want to know, <delay> <8> <OD> *that* anything much can now either <delay> <1> <MVB> *hurt* me or help me. (GB01:55)

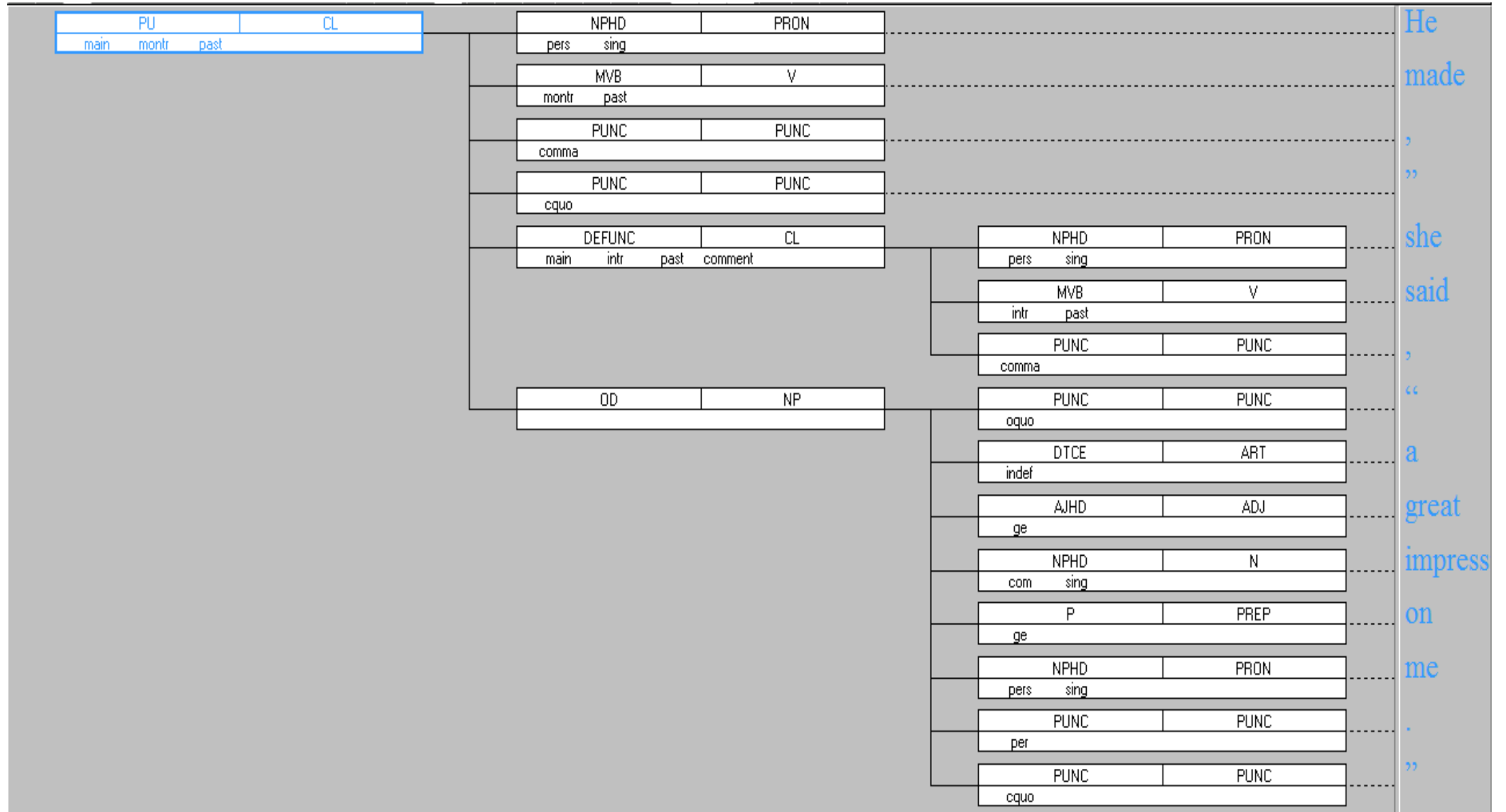
In contrast, example 21) offers a clear instance of an interruption between the auxiliary *don't* and the main verb *see*. (Note that *at least* is counted as one word, being one of the linked expressions described above.) However example 22) is more complex. *Hurt* is shown as being delayed by one word because *either* cannot go anywhere else but *now* could be placed, for example, at the end of the sentence.

In some cases, delay may be an example of what Jeffries (2010) calls 'syntactic iconicity'. Syntactic iconicity is the use of (possibly mildly) deviant syntax which evokes feelings in the reader which 'directly correspond to the emotions being evoked in the content' (Jeffries, 2010, p. 96) of a literary text. Jeffries suggests (personal communication) that this sentence from *Washington Square* may be an example of this technique.

- 23) He suffered a week to elapse, and then one day, in the morning, at an hour at which she rarely saw him, he strolled into the back parlour. (WS31:34)

At this point in the story, Morris Townsend has broken off his engagement to Catherine Sloper but Catherine's father does not know this. Catherine has been expecting her father to broach the subject of her engagement as her aunt has warned her that he will do so. The structure of this sentence, with its four adverbial phrases, one post-modified by a relative clause, delays the arrival of the second main clause which the reader has expected since the introduction of the coordinator 'and'. The reader may then feel some impatience or tension or anxiety about the coming encounter, because of this unusual delay, which is exactly what Catherine is feeling. Consciously or unconsciously, the syntax of the sentence might lead the reader to experience the emotion the text is describing. If so, there is 'a direct reflection of the dynamics of the situation in the structure of the syntax' (Jeffries, 2010, p. 113).

Figure 4-11 GB22:79 showing detached function and delay



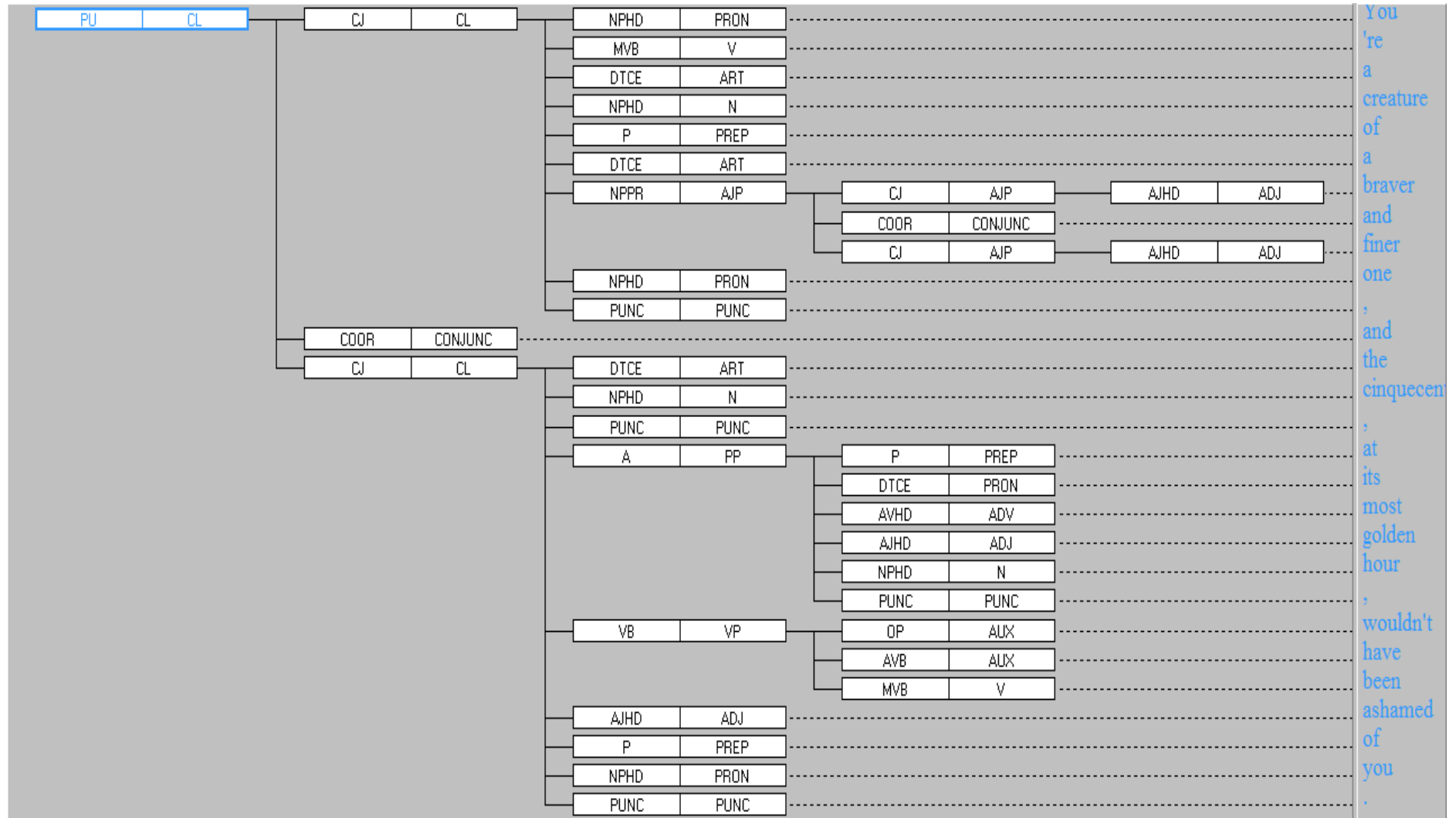
<#79:1> <sent> <ds> <D2> <O!> </O!> He made, she said, <delay> <?> <OD> “a great impression on me.”

#### 4.4.4 Coordination

Coordination is shown in ICECUP at the phrasal level. For this reason, I have inserted phrasal nodes when items are coordinated so that I can assess the amount of coordination in different parts of the corpus. Figure 4-12 shows a sentence from the first chapter of *The Golden Bowl* which contains coordinated clauses and adjectives (AJP), while Figure 4-13 shows a detail of the sentence, illustrating the feature *coordination* in the phrase node for the coordinated adjectives *braver* and *finer*, as well as on the whole parsing unit.

In Figure 4-13, the coordinated adjectives have extra phrase nodes which are added in a coordinated construction so that the coordinated items are linked together. Here they make up a noun phrase premodifier (NPPR), which has the form of an adjective phrase and is marked with the feature *coordination* (*coord*) in the lower part of the box as well as having their own adjective phrase nodes which are given the function *conjoin* (*CJ*).

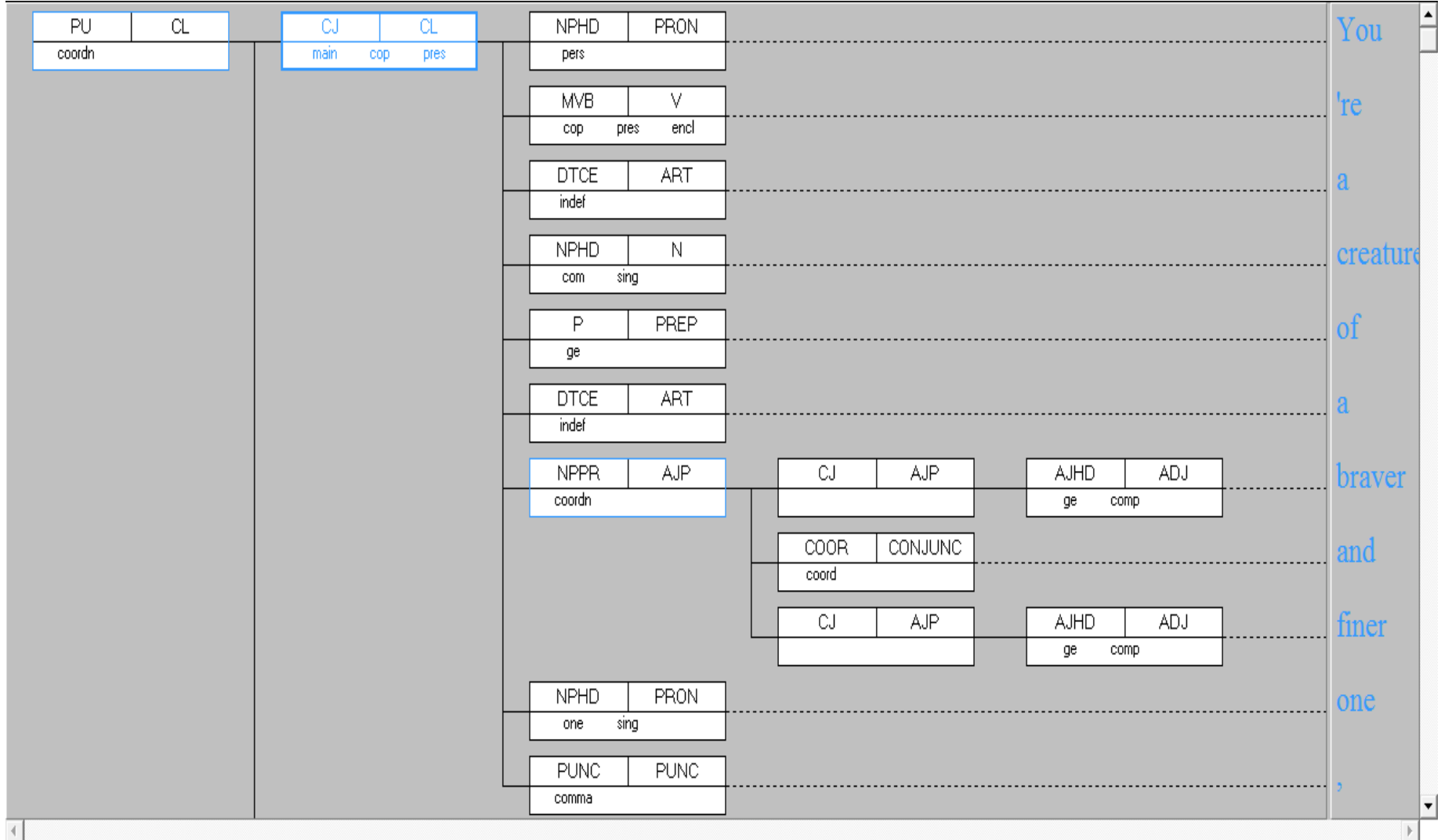
Figure 4-12 GB01:148 illustrating coordinated clauses and adjectives



You're a creature of a braver and finer one, and the *cinquecento*, at its most golden hour, wouldn't have been ashamed of you.



Figure 4-13 GB01:148 detail showing coordination feature at phrase level



You're a creature of a braver and finer one, and the *cinquecento*, at its most golden hour, wouldn't have been ashamed of you.

## 4.5 Hard cases

Many sentences present challenges to the parser (as well as the reader) in *Washington Square* and particularly in *The Golden Bowl*. In this section I will present various examples to illustrate some of the difficulties which have arisen in the process of compiling the *HJPC*.

### 4.5.1 Self-correction

When including direct speech in his novels, James sometimes uses the device of self-correction. For example,

24) "No," said Catherine; "I have asked him — asked him to wait." (WS18:37)

This presents a problem for my annotation system. I read this sentence as hesitant self-correction as Catherine is described as 'afraid' in the following sentence. In ICE-GB there are many examples of self-correction in the spoken part of the corpus. The ICECUP convention is to delete the speaker's first attempt (though the words remain visible) and to use the second attempt as the sentence which needs to be parsed. This is the convention followed in the *HJPC*, but all the words are counted when calculating the number of words in the sentence. The deleted words are marked in red, so the sentence appears in *WS* as:

25) "No," said Catherine; "I have asked him — <delay><2><MVB> asked him to wait." (WS18:37)

The second *asked* is shown in the tree diagram as the main verb in the sentence. This is counted as delay; WS18:37 has a delay score of 2, representing the words between *have* and the second *asked*.

### 4.5.2 What is not delay?

In general, adverbials cannot be delayed, as explained in section 4.4.3, because their position in a clause is, with some exceptions, flexible. Sometimes this appears anomalous. For example:

26) The Doctor was both puzzled and disappointed, but he solved his perplexity by saying to himself that his daughter simply misrepresented — justifiably, if one would, but nevertheless, misrepresented — the facts; and he eased off his *disappointment*, which was that of a man losing a chance for a little triumph that he had rather counted on, *by a few words* that he uttered aloud. (WS31:63)

In this sentence, the punctuation seems to indicate the prepositional phrase beginning *by a few words* has been delayed from the noun *disappointment*, but this

would not fit the rule that an adverbial cannot be delayed, so I have not counted delay here.

There is also no delay counted between a coordinator and the start of a coordinated item. In example 27) which shows WS31:34, a coordinated main clause has a number of adverbial phrases between the coordinator (*and*) and the subject of the clause (*he*). No delay is marked, in line with my treatment of any other main clause. Nevertheless, this sentence is undoubtedly deviant in that it differs from what Jeffries calls 'a notion of the linguistically normal' (Jeffries, 2010, p. 108) which a native English reader could be expected to have. This deviation could be confirmed by searching for similar structures in a natural language reference corpus such as ICE-GB.

- 27) He suffered a week to elapse, *and* then one day, in the morning, at an hour at which she rarely saw him, *he* strolled into the back-parlour. (WS31:34)

### 4.5.3 Phrasal verbs

Phrasal verbs illustrate a problem with trying to create a corpus which is completely consistent and follows rules which are easy to understand. I would usually count delay where an adverb intervenes between a verb and its complement, as shown in example 28). However with phrasal verbs there may be no other place for the adverb to be placed grammatically and, in such cases, delay does not apply. On the other hand, if the adverb could be placed after the complement, then delay does apply. For example:

- 28) <D1> And he got up, *holding* out <delay> <1> <DO>*his hands* toward her. (WS18:41)
- 29) There had been, through life, as we know, few quarters in which the Princess's fancy could let itself loose; but it *shook* off *restraint* when it plunged into the figured void of the detail of that relation. (GB38:28)

In example 29), however, delay is not marked between *shook* and the direct object *restraint* because it seems to me that it would be odd, though not ungrammatical, if the sentence read *it shook restraint off when it [...]*. This is a subjective judgment with which others might well disagree, and, where there are examples of this type, the corpus is inevitably subjectively analysed rather than completely rule-driven.

This classification of phrasal verbs follows the Quirk, et al., (1985) grammar. It would be possible to identify phrasal verbs separately by annotation and to decide that their associated adverbs (or particles) were part of the verb and therefore did not occasion delay. However, this would not capture the distinction between example 28) above and the possible alternative:

And he got up, *holding his hands* out toward her.

As I wish to capture examples of delay when they occur, I have not made any special classification for phrasal verbs.

#### 4.5.4 Apposition

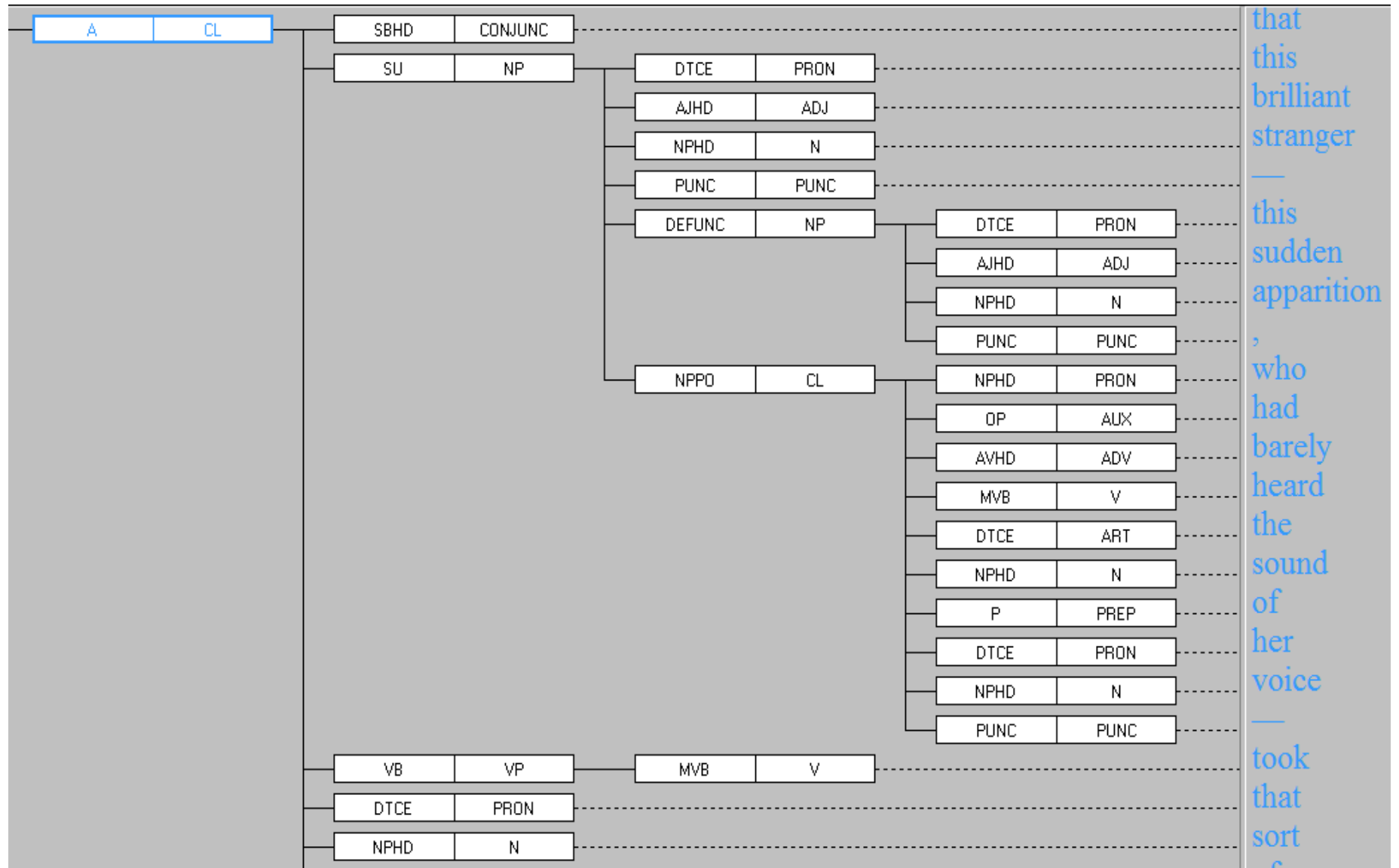
As explained in section 4.4.3, in ICECUP apposition is dealt with by marking the second element as a detached function (DEFUNC) with the feature *apposition*. This is illustrated with sentence WS05:125, given below as example 30), with a view of the relevant section only in the tree diagram in Figure 4-14. (This sentence also contains an APU.)

- 30) <D15>] But that **this brilliant stranger** — this sudden apparition, <delay> <3> <NPPO> who had barely heard the sound of her voice — <delay> <12> <VB> **took** that sort of interest in her that was expressed by the romantic phrase of which Mrs. Penniman had just made use: <apu> this could only be a figment of the restless brain of Aunt Lavinia, whom every one knew to be a woman of powerful imagination. (WS05:125)<sup>38</sup>

---

<sup>38</sup> This sentence is from Chapter 5 of *Washington Square*, which has been parsed but not included in the *HJPC* as lack of time prevented the inclusion of a matching chapter from *The Golden Bowl*.

Figure 4-14 Part of WS05:125 illustrating apposition



## **4.6 Conclusion**

In compiling this corpus, some constraints and limitations have inevitably been encountered. The processing of the sample chapters has been extremely onerous and that has prevented me from including more data, which I would have preferred. I have striven to produce the most consistent and objective database possible, but errors are unavoidable. To the extent that these are random and relatively unusual, the data which is collected from the corpus should not be biased by the presence of errors. The corpus remains subjective insofar as, where the reading of a sentence is ambiguous or where a decision has to be made about grammatically acceptable alternative phrasing (see section 4.5.3), the analysis reflects my personal interpretation and idiolect. This has been unavoidable as I have worked on it alone.

With these caveats, the process described in this chapter has allowed me to produce a corpus which can be analysed to research my central question of the characteristics of the 'difficult late style' compared with an early example, at least in these two novels. This process and the results of the analysis will be set out in Chapter 5.

## Chapter 5      Analysing the Data

---

Having compiled the *HJPC* (*Henry James Parsed Corpus*) from samples of *Washington Square* and *The Golden Bowl* as described in Chapter 4, the data was then explored and analysed using the search capabilities of ICECUP. There are potentially an enormous number of different questions which could be addressed using the information which the corpus contains, but this dissertation concentrates on the exploration of the characteristics of the syntax of James's late style and any contrast which exists with his early style. Starting from simple statistics on the two samples, this chapter will discuss the question of what 'difficulty' might mean in relation to James's syntax and to what extent this relates more to *GB* than *WS*. Although the general focus will be on the data obtained from the whole of *WS* and *GB*, individual exceptional sentences will also be discussed.

In the analysis of the data produced from the *HJPC*, a Wilson interval test was used to calculate confidence intervals for the data provided, illustrating these with error bars on the graphs. All results in this dissertation are calculated to a 95% confidence interval. This means that there is a 95% probability that the observations are statistically significant rather than being found by chance. The standard test to calculate confidence intervals is the 4x4 (or chi-squared) test which assumes that the data under analysis is distributed on a normal curve. However, created artefacts like novels do not follow a normal distribution but tend to follow a Poisson distribution. A test of data from the *HJPC* showed that this was indeed the case. In these circumstances, the Wilson test, the standard test for confidence intervals for a Poisson distribution, will provide more accurate confidence results. In essence, the Wilson interval test applies a mathematical algorithm to a Poisson distribution to convert it to a normal distribution, and then applies the chi-square test to it.

Where the confidence intervals are small and do not overlap, the test shows that the results are statistically significant and can be taken to represent differences between the whole of the two novels under discussion, *Washington Square* and *The Golden Bowl*, rather than just the samples in the *HJPC*. The size of the *HJPC* is sufficiently large to be able to draw robust statistical conclusions for the two novels as a whole.<sup>39</sup>

(The colour scheme and layout shown in the graphs is consistent throughout, with *WS* always shown on the left in turquoise and *GB* on the right in navy.)

---

<sup>39</sup> For a fuller explanation of the use of the Wilson test and an example spreadsheet, see Sean Wallis's article at <http://corplingstats.wordpress.com/2014/02/05/binomial-normal-wilson/#more-1162>

## 5.1 Vital statistics

Even before using ICECUP, initial data can be gleaned from the e-texts themselves. ICECUP does not provide a word count facility, so it was necessary to transfer each chapter to Microsoft Word, which provides an automatic count. However, the definition of a 'word' for this count is not the same as that used by ICECUP, where two words can be marked as one (ditto-tagged), for semantic and functional reasons, as described in Section 4.4.3. For example, in ICECUP *United States* is treated as one word when it refers to the single nation whereas Microsoft Word will count them as two. ICECUP does, however, organise the text into sentences which it numbers, and a simple search reveals the number of clauses in a text. These searches yield the figures shown in Table 5-1.

**Table 5-1 Vital statistics of the HJPC**

Source	Words	Sentences	Clauses
<b>WS</b>	8,888	482	1,567
<b>GB</b>	23,999	993	3,568

Even a physical examination of the two novels makes it clear that *The Golden Bowl* is a far longer novel than *Washington Square*. *The Golden Bowl* has more chapters and longer chapters than *Washington Square*, so, even though the HJPC includes the same number of chapters from each book, the amount of data this provides from *The Golden Bowl* is much greater than that from *Washington Square*, as shown in Table 5-1. The same figures are used to generate Figure 5-1.



**Figure 5-1 Vital statistics compared**

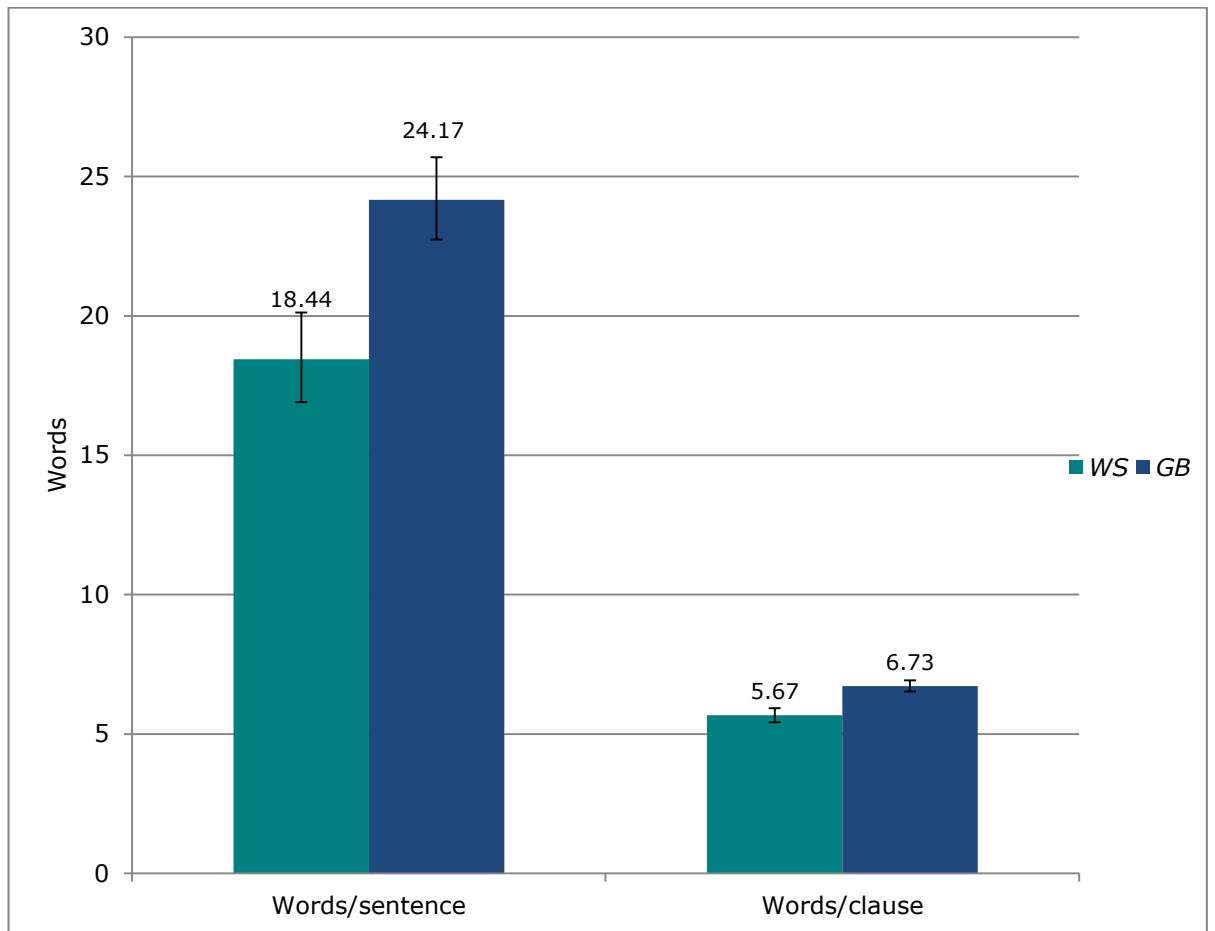


Figure 5-1 shows in graphic form that the average sentence length in GB is substantially greater than that in *WS*. However, the difference in clause length is much smaller. It is statistically significant but does not explain how the greater sentence length of *GB* is achieved. This is clarified in Table 5-2 and the explanation which follows it.

**Table 5-2 Ratio of *WS* to *GB***

	<b>Ratios <i>WS</i> : <i>GB</i></b>
<b>Words</b>	1 : 2.7
<b>Sentences</b>	1 : 2.06
<b>Words per sentence</b>	1 : 1.311
<b>Clauses</b>	1 : 2.277
<b>Words per clause</b>	1 : 1.19

Table 5-2 shows the relationship of *WS* to *GB* in terms of the number of words, sentences and clauses they contain, and the length of sentences and clauses. In terms of words, *GB* is 2.7 times larger than *WS*. The sentence ratio is smaller at 1: 2.06 because *GB*'s sentences are longer by a factor of 1: 1.3. However, as Figure 5-1 showed, this is not mainly because *GB* has longer clauses; the ratio of words per clause is close at 1: 1.19. Rather, there are many more clauses in *GB* with a ratio of 1: 2.277. Thus the difference between *WS* and *GB* is that *GB* has longer sentences, due in small part to slightly longer clauses but to a much larger extent to the greater number of clauses. This is the first indication of the characteristic syntax of James's late style, which involves more elaborately structured sentences.

Section 5.2 gives the result of an analysis of the type of clauses which make up these longer sentences, after first discussing the relationship of the clause types found with syntactic complexity and the 'difficulty' of James's late style.

## 5.2 Difficulty and complexity

When critics such as Chatman (1972) have used the word 'difficult' in describing James's late style,<sup>40</sup> they are referring to the reader's experience of James's novels, in which they might have actual difficulty in understanding his prose, or find it difficult to continue to read or enjoy his work because of the language he uses. (Other difficulties with reading James might arise from factors unrelated to the language he uses, such as the remote period or social context in which the novels are set, or ambiguity in the attitude of the protagonists or their situation.) The notion of difficulty does not correlate exactly with any particular characteristic of syntax. However, sentences can be described as syntactically complex, a concept which contrasts with compound sentences. Complex sentences contain one or more dependent clauses whereas compound sentences contain coordinated clauses. The two definitions are not mutually exclusive – a sentence can be both compound and complex. It is this idea of complexity that is initially explored here, with the hypothesis that syntactic complexity is likely to be a contributor to reader-perceived 'difficulty'.

### 5.2.1 Other measures of complexity

Other measures of complexity have been suggested. Biber, Conrad and Reppen (1998, p. 178) refer to the idea of a T-unit, – 'an independent clause with all its dependent clauses' - which is used in reference to children's writing, and whose length can be measured and comparisons made. A similar approach is taken by Hallen and Shakespear (2002), who use the T-unit as an aid to interpreting Emily

---

<sup>40</sup> See section 3.2.6.

Dickinson's poems, which they identify as difficult to read. They also suggest the use of the T-unit as a measure of complexity:

T-unit length can serve as a measure of syntactic complexity and cognitive maturity in a writer or author. Less-experienced writers tend to conjoin shorter sentences in a paratactic style, while skilled writers tend to combine sentences into longer main clauses with subordinate dependent clauses in a hypotactic style. (Hallen & Shakespear, 2002, p. 91)<sup>41</sup>

Hallen and Shakespear track the length of T-unit used by Dickinson in her poems from the early, middle and late periods of her life and conclude that Dickinson used longer T-units, i.e. greater syntactic complexity, as her career progressed with a slight dip towards the end of her life.

McNamara, Louwerse, Cai and Graesser have developed the Coh-Metrix tool, which is held at the University of Memphis (McNamara, et al., 2005). An online 'Text Easability Assessor' is provided which scores a short passage of text for readability on five components: 'word concreteness, syntactic simplicity, referential cohesion, deep cohesion, and narrativity'. This is based on an underlying analysis of 108 indices from word, sentence and paragraph counts to measures such as the number of modifiers per noun phrase.<sup>42</sup> These indices are grouped into eleven different text characteristics, one of which is 'syntactic complexity', which is relevant to this project (although the other factors they analyse may also affect a reader's perception of 'difficulty'). The factors which contribute to the component 'syntactic complexity' include measures of the amount of modification in a noun phrase, the length of the sentence before the main verb, and the similarity of the sentence to other sentences. All the measures assess syntactic complexity indirectly rather than actually parsing the text. Using this tool, they analyse text complexity, stating, for example, that 'all five dimensions reveal that *Little Women* is less complex than *The Great Gatsby*' (Graesser, et al.).<sup>43</sup> This is a different definition of syntactic complexity to the syntactic definition used in this dissertation; the latter is simply a term for sentences containing dependent clauses, as explained in section 5.2.

Szmrecsanyi (2004) discusses the history of the idea of syntactic complexity, which he says has been taken to be indicated mainly by the number of words in a sentence, but can also be counted as the number of words in a node (or branch of a syntactic tree), or as a contrast between coordination and subordination. Further possibilities include counting other items such as subordinators, non-finite verbs, etc. Szmrecsanyi experiments with a small 50 sentence corpus, tagged and parsed, including written and spoken text. He tries out

---

<sup>41</sup> Hallen and Shakespear cite Hunt, 1970 in support of this claim.

<sup>42</sup> For the complete list, see the McNamara et al. (2005) website.

<sup>43</sup> Further details of the analysis used by Coh-Metrix is given in Graesser, et al., 2004

the different methods he has discussed, simplifying them into three measures: sentence length, a combined index of syntactic complexity, and node count. He finds 'that the three structural measures rank the data in an indeed surprisingly conform fashion' (Szmrecsanyi, 2004, p. 1036).

Givón defines syntactic complexity in terms of an increase in hierarchy.

The gist of Herbert Simon's (1962) seminal work on the architecture of complexity is that increased complexity is, at the most general level, increased **hierarchic organization**; that is, an increase in the number of hierarchic levels within a system [Bold emphasis is Givón's] (Givón, 2009, p. 4).

He suggests a logical relationship between the complexity of ideas to be communicated and complexity of syntax in a text, as follows:

- a. **Coding**: More complex mentally-represented events are coded by more complex linguistic/syntactic structures.
- b. **Processing-I**: More complex mentally-represented events require more complex mental processing operations. Therefore,
- c. **Processing-II**: More complex syntactic structures require more complex mental processing operations. (Givón, 2009, p. 12)

However, he admits that the experimental evidence for these assumptions is still uncertain and contradictory.

## 5.2.2 Coordination and dependency

I will introduce a more detailed measure of complexity in the next chapter, taking up some of the ideas discussed by Szmrecsanyi. Here, following Givón, a measure of the coordinated and dependent clauses in each novel gives a first picture of the relative syntactic complexity of the two books. Many dependent clauses in a sentence often lead to a hierarchic organization, which Givón suggests requires more mental effort and that may be subjectively interpreted as 'difficult'. An example is shown in Figure 5-2, which shows part of WS01:12. The text of this example is:

- 31) putting aside the great good-nature of the circle in which he practised, which was rather fond of boasting that it possessed the "brightest" doctor in the country (Part of WS01:12)

The adverbial clause which begins with *putting* is inside a coordinated clause which is already one step lower than the overarching main clause. Within the adverbial clause are two clauses at the same level both postmodifying the noun *circle* – *in which he practised* and *which was rather fond of boasting* - and two more clauses beginning with *boasting* and *that* which occupy successively lower places in the hierarchy. This example from *Washington Square* has many levels of hierarchy within its syntax.

An analysis of the quantities of main, coordinated and dependent clauses in the two parts of the *HJPC* allows a comparison of the hierarchical structures of the syntax of the early and late novels. This also echoes Cross's (1993) suggestion that a comparison of ratios of coordination and dependency can serve as a method of discrimination between the styles of *WS* and *GB* (see section 3.2.5). Coordinated clauses are shown in ICECUP at a lower level of the tree hierarchy with the function *Conjoin* (CJ) and are branches of a dominating clause which has the feature *coordination* (coord). Such clauses can be identified with a *fuzzy tree fragment*, a special search facility included in ICECUP. This allows the user to construct part of a tree diagram using blank nodes which can be completed to a greater or lesser extent, making the fragment more or less 'fuzzy'. Once the search has been done, ICECUP will show the number of coordinated clause examples it has found in each sentence, the total for the data searched, as well as the number of sentences they were found in. An example of a coordinated sentence which is found by such a search is shown in Figure 5-3. This short compound sentence shows the dominating clause, which here has the function of a Parsing Unit, highlighted in blue on the top left. The coordination feature is in the lower part of the node, as explained in section 4.2. The two coordinated clauses on the next level of the tree, highlighted in red, are joined by the coordinating conjunction *and* with the components of each clause shown on the right.

Similarly, dependent clauses can be identified because they carry the feature *dependent* (depend). Once again they are identified using a fuzzy tree fragment, chapter by chapter. Two different examples are shown in Figure 5-4, where the head node of each clause is marked in green. Note that Figure 5-4 shows a fragment of a sentence rather than the full tree. The first dependent clause shown postmodifies the word *things* and is a relative clause, whereas the second is an adverbial non-finite clause.

Figure 5-2 Part of WS01:12 illustrating clausal hierarchy

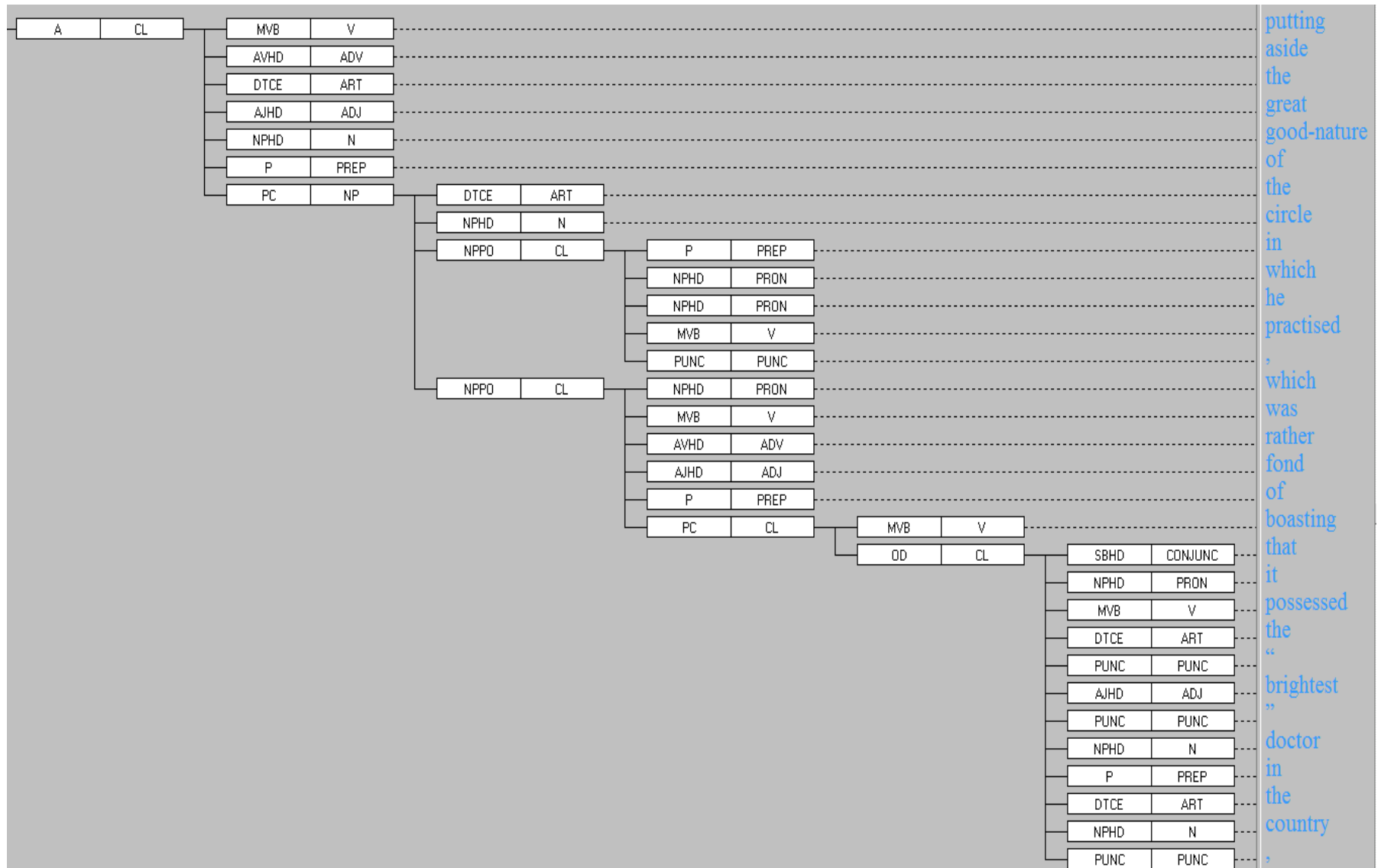


Figure 5-3 WS18:57 showing coordinated clauses

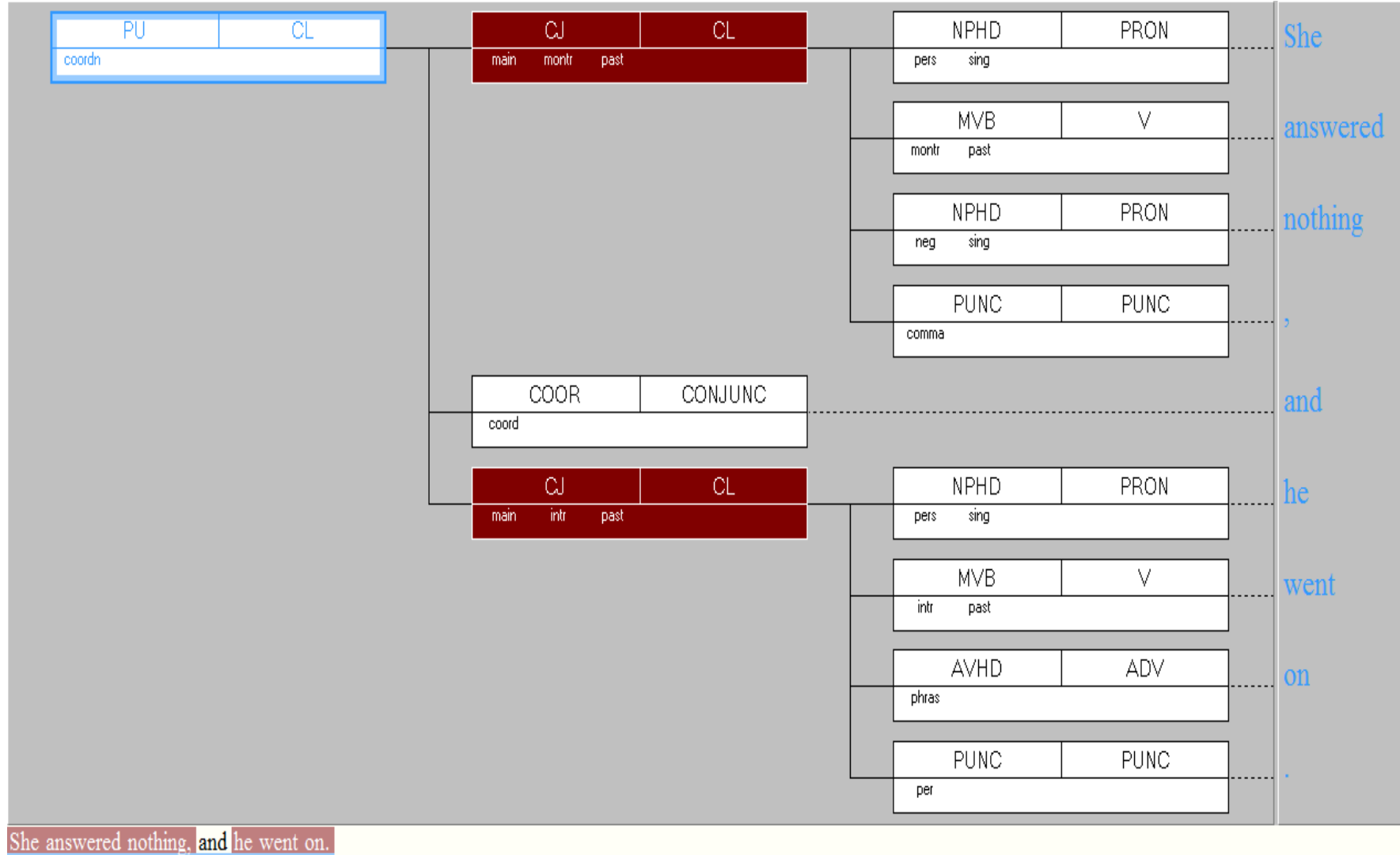
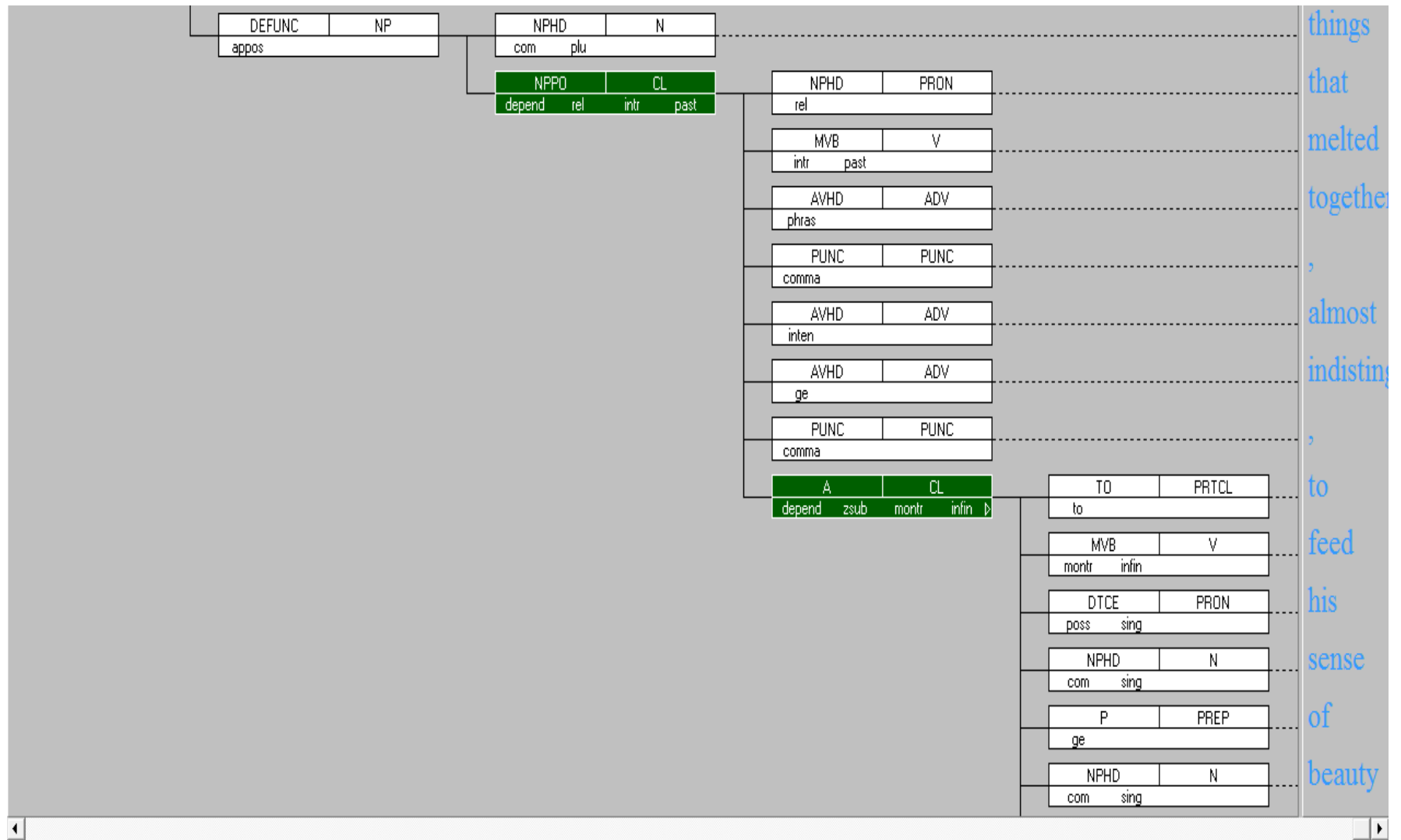


Figure 5-4 Part of GB22:9 illustrating dependent clauses



But there were more things before him than even these; things that melted together, almost indistinguishably, to feed his sense of beauty.



The two parts of the *HJPC* were searched, using the methods described above, chapter by chapter, eliciting data for each one and then calculating totals. Figure 5-5 displays the results of these searches, comparing the number of main, coordinated and dependent clauses in *WS* and *GB*. Some dependent clauses are also coordinated, so there will be some overlap between these categories, and some clauses are neither main, coordinated nor dependent (see the overarching clause in Figure 5-3), but these exceptional clauses are not included in the statistics for Figure 5-5.

**Figure 5-5 Coordinated and dependent clauses**

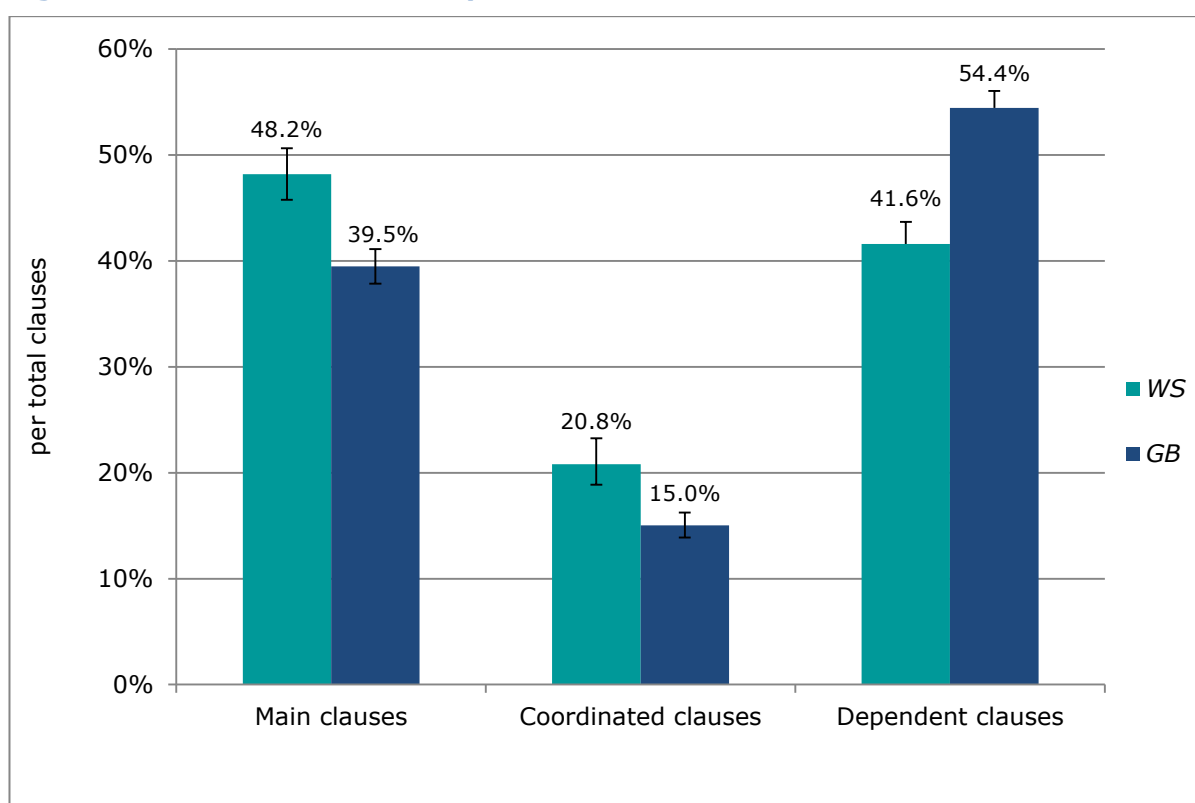


Figure 5-5 shows that, while the overall amount of clauses per sentence and clausal length is similar in *WS* and *GB* as seen in Figure 5-1, the type of clauses used by James in the two books is different and to a statistically significant extent. In *WS* James uses more main clauses than dependent clauses (48.2% main as compared to 41.6 % dependent) whereas in *GB* the reverse is true; there are more dependent clauses (39.5% main as compared to 54.4% dependent). The ratio of coordinated to dependent clauses in *WS* is 1:2, while in *GB* it is 1:3.6. This confirms the hypothesis that, if a higher level of dependency is considered a marker of complexity, *The Golden Bowl* is more complex than *Washington Square*. Figure 5-5 also shows that Cross's (1993) suggestion that 'compounding' is an important element of James's early style (see section 3.2.5) has a basis in this comparison of *Washington Square* and *The Golden Bowl*. There are considerably more coordinated

clauses in the early novel. The ratios are 2.34:1 main to coordinated clauses for *WS* and 2.63:1 for *GB*.

While a greater use of dependent clauses has been shown to be one characteristic of the late style, as well as a lower use of coordinated clauses, this analysis has not yet included the question of discourse type. If one of the samples contains more direct speech, and if James uses a different style for direct speech, that might affect the figures shown in Figure 5-5. This question is explored in Section 5.2.3.

### 5.2.3 Complexity and the representation of speech

Greenbaum and Nelson (1995, p. 2) have pointed out that 'there is [...] a widespread view that the spoken language is characterized by parataxis (coordination and juxtaposition of clauses) and the written language by hypotaxis (subordination of clauses)'. They set out to investigate this view in their article by means of a corpus analysis of a selection of written and spoken texts from ICE-GB, and their conclusions generally support the hypothesis, although the results are partly dependent upon the formality of the written data type. However, I suggest that the existence of this widespread view in itself is likely to make novelists represent direct speech using coordination rather than subordination. If this is the case with James, then the proportion of direct speech in the two parts of *HJPC* might affect the proportion of coordinated and dependent clauses shown in Figure 5-5.

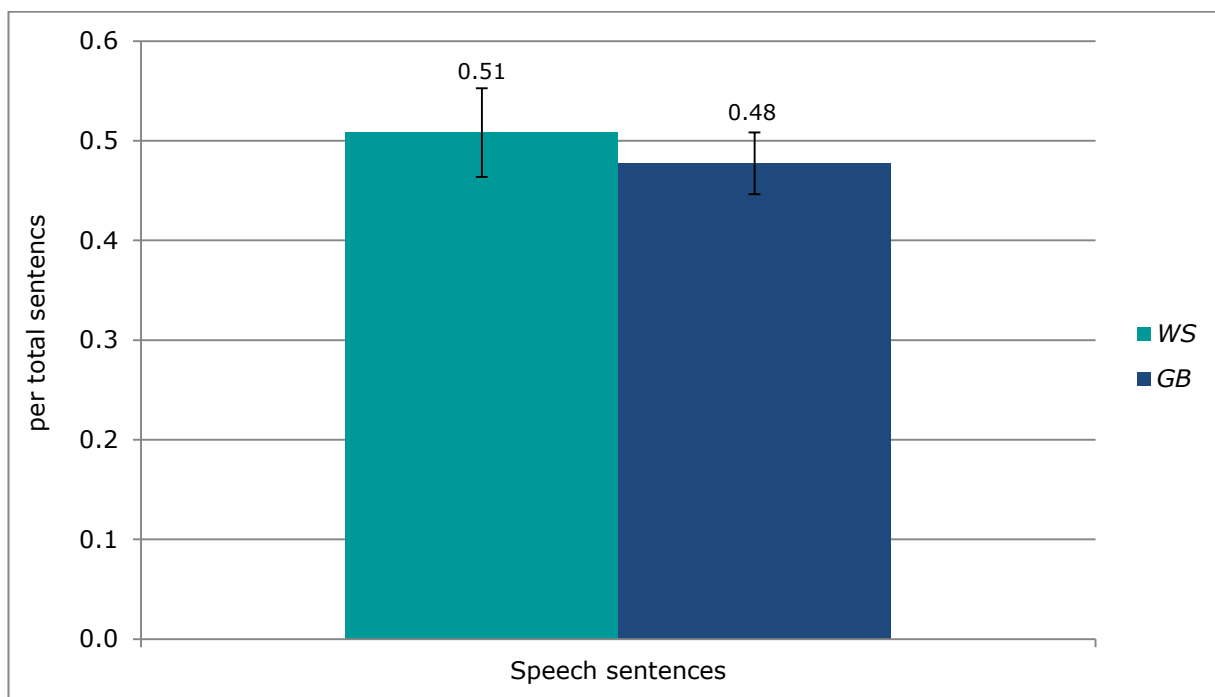
Therefore it is useful to separate the speech sentences (those with the markup code <ds>, showing that they contain speech marks<sup>44</sup>) from non-speech sentences to see how much there is of each type of discourse in the two novel samples. Figures for the speech sentences within each chapter were obtained and totals calculated to produce the data used for Figure 5-6. The graph shows that there is a slightly larger proportion of speech sentences in *WS* than in *GB*, a ratio of 51% to 48% of the total number of sentences. However, this result is not statistically significant and the two novels may be considered to have similar ratio of speech to total sentences.

*Washington Square* and *The Golden Bowl* do, however, have different ratios for the number of clauses per sentence. Figure 5-7 shows the proportion of speech clauses in the two parts of the corpus, and shows a greater differentiation with approximately 10% more speech clauses in *WS* as compared to *GB*.

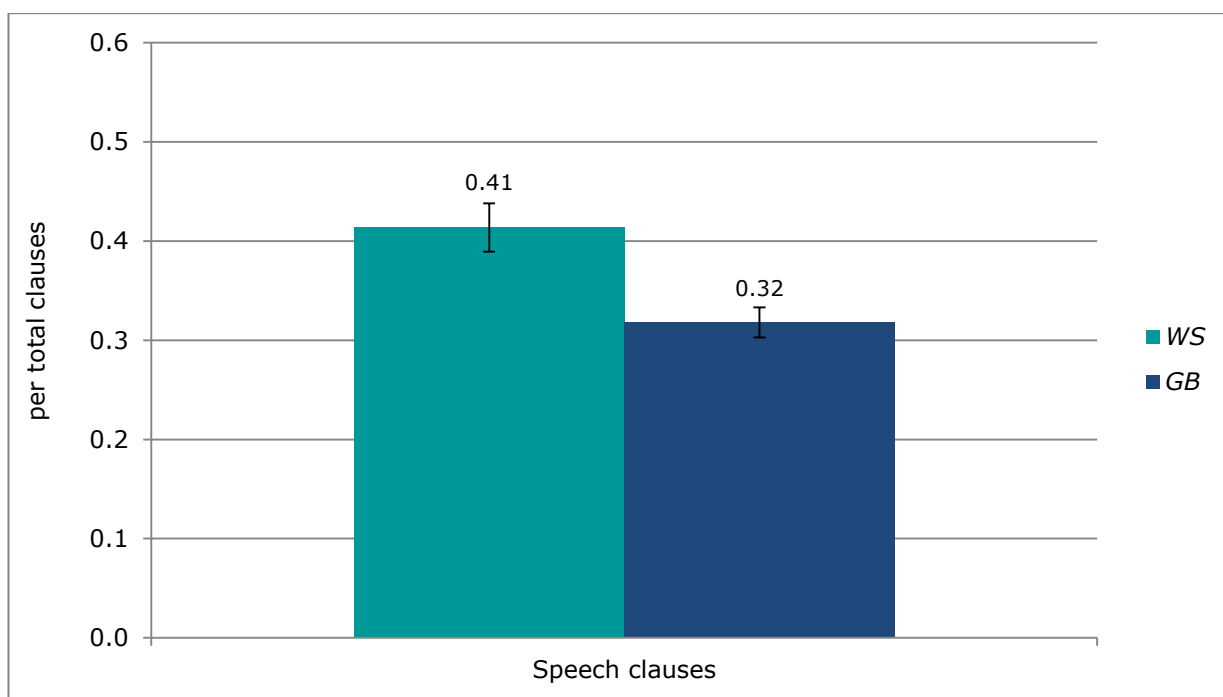
---

<sup>44</sup> See section 4.4.1.

**Figure 5-6 Relative proportion of speech and non-speech sentences**



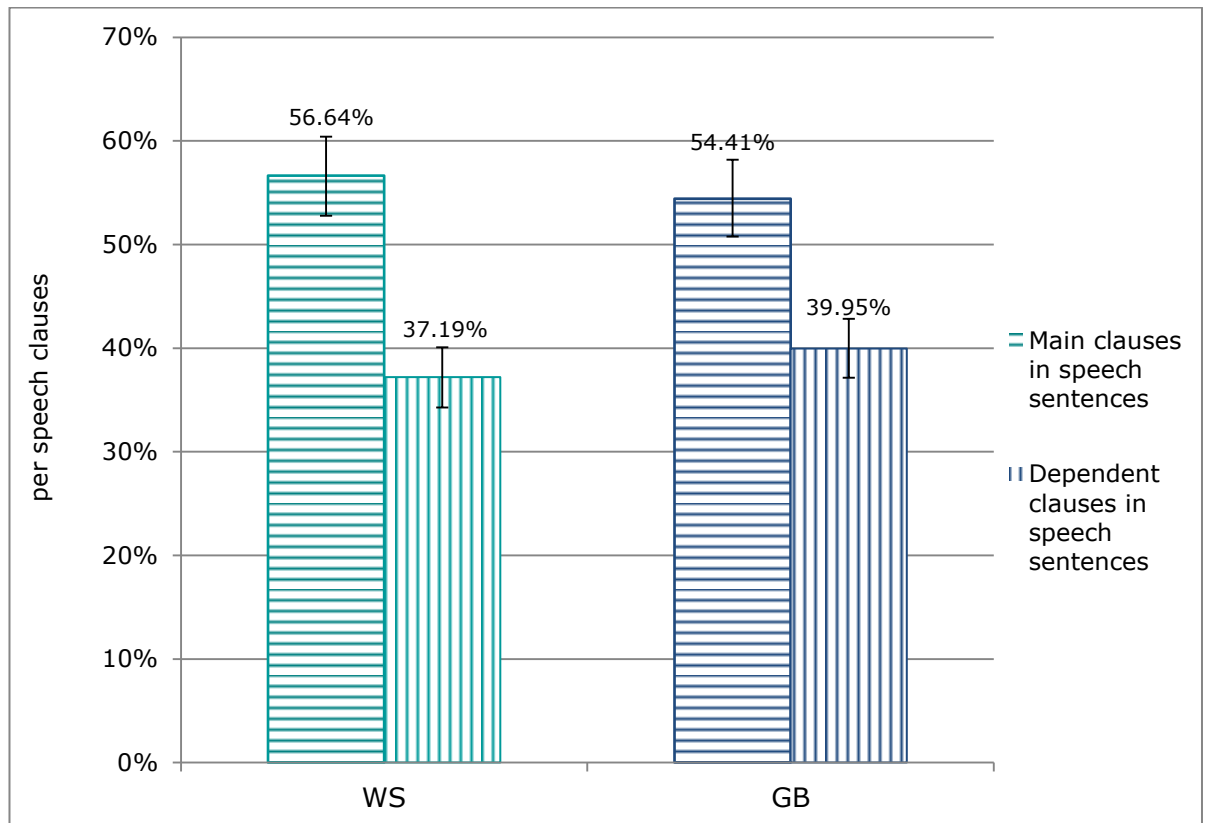
**Figure 5-7 Relative proportion of speech and non-speech clauses**



As *WS* has more speech clauses than *GB*, it is now necessary to discover whether speech sentences in the *HJPC* have fewer dependent clauses than non-speech sentences. If that is the case, then that might partly or completely explain the increased percentage of dependent clauses in *GB* shown in Figure 5-5, rather than that graph demonstrating a genuine difference between James's early and late styles in these novels. Here the ICECUP query facilities are again used to find

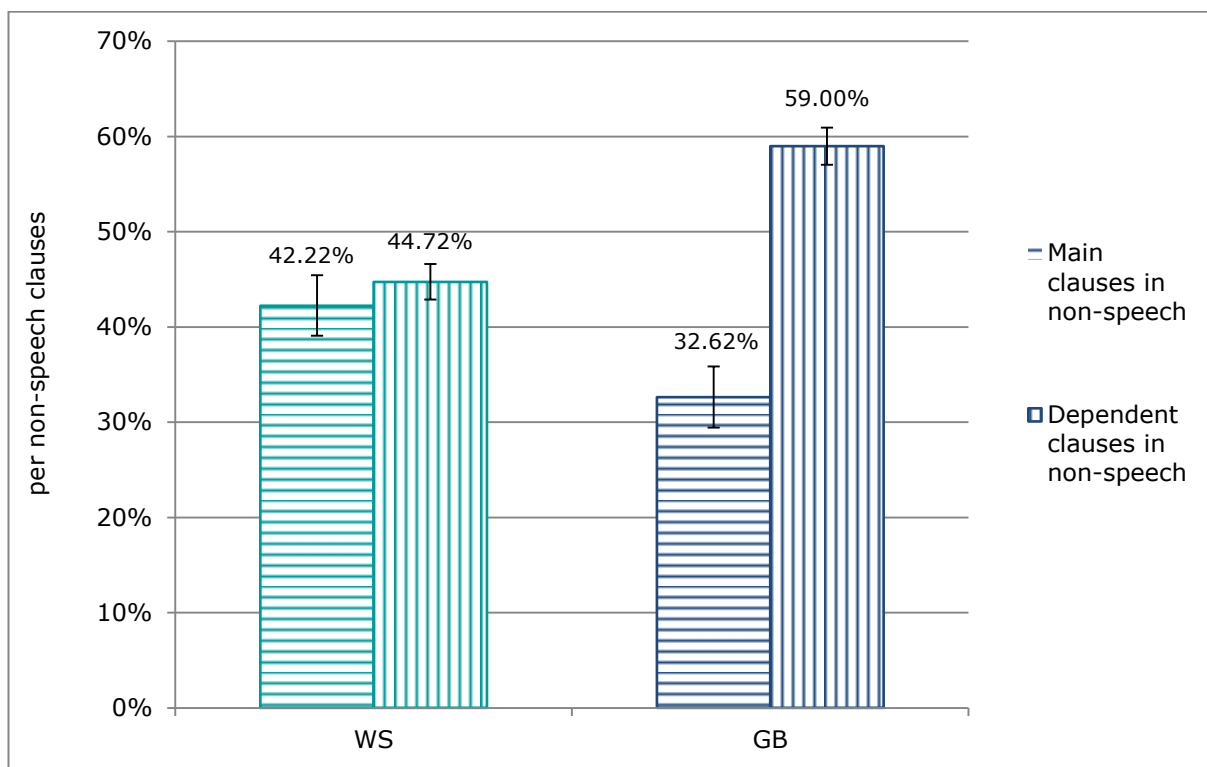
dependent clauses, but within the previous search which has found the <ds> speech sentences. The figures for non-speech sentences are found by simply deducting the speech clause figures from the overall number of dependent clauses. Error bars were then added to show the confidence intervals calculated for these figures using the Wilson test. Figure 5-8 shows the result and it is clear that, within speech sentences, a far smaller proportion of dependent clauses are used than in the texts overall. In *WS* 37.19% of clauses are dependent in speech sentences whereas the proportion for all sentences is 41.6% (as shown in Figure 5-5). Similarly in *GB* there is a decrease, with 39.95% of clauses being dependent compared to 54.4% overall. Figure 5-8 also shows that the ratio of main to dependent clauses in *WS* and *GB* is extremely similar; the ratio figures are 1:1.5 for *WS* and 1:1.4 for *GB*. Clearly there is no specially marked late style in the amount of dependency used by James within speech sentences.

**Figure 5-8 Main and dependent clauses in speech sentences**



As we have previously seen (in Figure 5-5) that there is a definite increase in the proportion of dependent clauses in *GB* as compared to *WS*, that difference must be in the non-speech sentences. Figure 5-9 shows that to be the case.

**Figure 5-9 Main and dependent clauses in non-speech**



The comparison between the main and dependent clauses in non-speech sentences in *WS* is within the margin of error: the proportions are effectively the same. However, in *GB* there is a strong contrast between the use of main and dependent clauses in non-speech sentences; the ratio is 1:1.8. Figure 5-9 also shows the difference between the amount of dependency in non-speech in *WS* and *GB*. 59% of non-speech clauses are dependent in *GB* while only 45% are in *WS*. So James's late style is more syntactically complex than his early style (i.e. uses more dependency) only in non-speech sentences. The 'difficulty' which readers report is likely, at least in part, to be related to the complex relational structures which a large number of dependent clauses convey. For example, in this sentence from the end of *The Golden Bowl*, the reader has to negotiate a complicated architecture of dependent clauses describing the time and place and behaviour of Maggie set within the (also complex) idea of her seeing the dawn connected to her understanding of her husband's behaviour.

- 32) It had never occurred to his wife to pronounce him ingenuous, but there came at last a high dim August dawn when she couldn't sleep and when, creeping restlessly about and breathing at her window the coolness of wooded acres, she found the faint flush of the east march with the perception of that other almost equal prodigy. (GB38:107)

It is thus possible to conclude that the style of *The Golden Bowl* differs from that of *Washington Square* by having more, though not longer, dependent clauses in sentences which do not contain direct speech.

#### 5.2.4 Exceptional sentences

Within this overall view, I suggest that particular exceptional sentences which have unusually high numbers of dependent or coordinated clauses make a disproportionate impact on the reader and their assessment of the style of the novel. This explains the judgment of critics such as Mizener (1966), Levin (1986), Cross (1993) and Fowler (1996), who discuss the difficulties of the late style as though it were homogenous (see Chapter 3). However, Springer (1993) does acknowledge the variable quality of James's writing (see section 3.2.5). Sentences of exceptional syntactic complexity can be identified by using ICECUP to provide a quantified list of dependent clauses and then noting by hand the number of dependent clauses per sentence, locating the sentences which have the highest number. Similarly coordinated clauses can be quantified in each sentence and located.

Table 5-3 shows the sentences with the highest number of coordinated and dependent clauses in each chapter, as well as the average for each set. As has been shown, there are more coordinated clauses in *WS* than in *GB*. 20.8% of total clauses are coordinated in *WS* but only 15% in *GB* (see Figure 5-5). These coordinated clauses are quite similarly distributed in sentences in both parts of the *HJPC*. The most coordinated sentence in *WS* is *WS18:4* with six coordinated clauses while in *GB* it is *GB20:10* with eight. Comparing the averages of the two sets, while *GB* is definitely higher, there has only been a small change from the early to the late sample. The picture is very different when exceptional sentences with the highest number of dependent clauses are compared. The difference between the lowest and highest score in *WS* is only two clauses whereas in *GB* it is five, and the average score is approximately double in the late sample as compared to the early one. There is a very stark contrast between *WS31:27*, the highest scoring *WS* sentence with seven dependent clauses, and *GB38:16*, the highest scoring *GB* sentence with thirteen, as can be seen by reading the sentence texts (examples 35) and 36) below). The full text of the highest scoring sentence in each category is given following Table 5-3.

**Table 5-3 Exceptional sentences**

Sentence number	Highest no. of coordinated clauses per sentence	Sentence number	Highest no. of dependent clauses per sentence
WS01:16	4	WS01:6	6
WS02:20	4	WS02:70	5
WS16:13	2	WS16:83	5
WS18:4	6	WS18:107	6
WS31:28	4	WS31:27	7
<b>Average</b>	<b>4</b>	<b>Average</b>	<b>5.8</b>
GB01:192	5	GB01:264	11
GB02:180	6	GB02:167	10
GB20:10	8	GB20:17	11
GB22:13	4	GB22:5	8
GB38:54	6	GB38:16	13
<b>Average</b>	<b>5.8</b>	<b>Average</b>	<b>10.6</b>

**Exceptional compound sentences**

- 33) She had an immense respect for her father, and she felt that to displease him would be a misdemeanour analogous to an act of profanity in a great temple; but her purpose had slowly ripened, and she believed that her prayers had purified it of its violence. (WS18:4)

[The start of each coordinated clause is indicated by the underlined word. In this sentence there are two overarching coordinated clauses which contain the first two coordinations and the second two coordinations, respectively, giving a total of six coordinated clauses in the sentence.]

- 34) He had paid first and last many an English country visit; he had learned even from of old to do the English things and to do them all sufficiently in the English way; if he didn't always enjoy them madly he enjoyed them at any rate as much, to all appearance, as the good people who had in the night of time unanimously invented them and who still, in the prolonged afternoon of their good faith, unanimously, even if a trifle automatically, practised them; yet with it all he had never so much as during such sojourns the trick of a certain detached, the amusement of a certain inward critical, life; the determined need, while apparently all participant, of returning upon himself, of backing noiselessly in, far in again, and rejoining there, as it were, that part of his mind that was not engaged at the front. (GB20:10)

[The start of each coordinated clause is indicated by the underlined word. This sentence also contains four conjoined noun phrases.]

### Exceptional complex sentences

- 35) It was an explanatory document, and it explained a great many things, chief among which were the considerations that had led the writer to take advantage of an urgent "professional" absence to try and banish from his mind the image of one whose path he had crossed only to scatter it with ruins. (WS31:27)
- 36) Maggie lived over again the minutes in question -- had found herself repeatedly doing so; to the degree that the whole evening hung together, to her aftersense, as a thing appointed by some occult power that had dealt with her, that had for instance animated the four with just the right restlessness too, had decreed and directed and exactly timed it in them, making their game of bridge -- however abysmal a face it had worn for her -- give way precisely to their common unavowed impulse to find out, to emulate Charlotte's impatience; a preoccupation, this latter, attached detectedly to the member of the party who was roaming in her queerness and was, for all their simulated blindness, not roaming unnoted. (GB38:16)

[In examples 35) and 36) the start of each dependent clause is indicated by the underlined word.]

### 5.2.5 Distribution of dependency

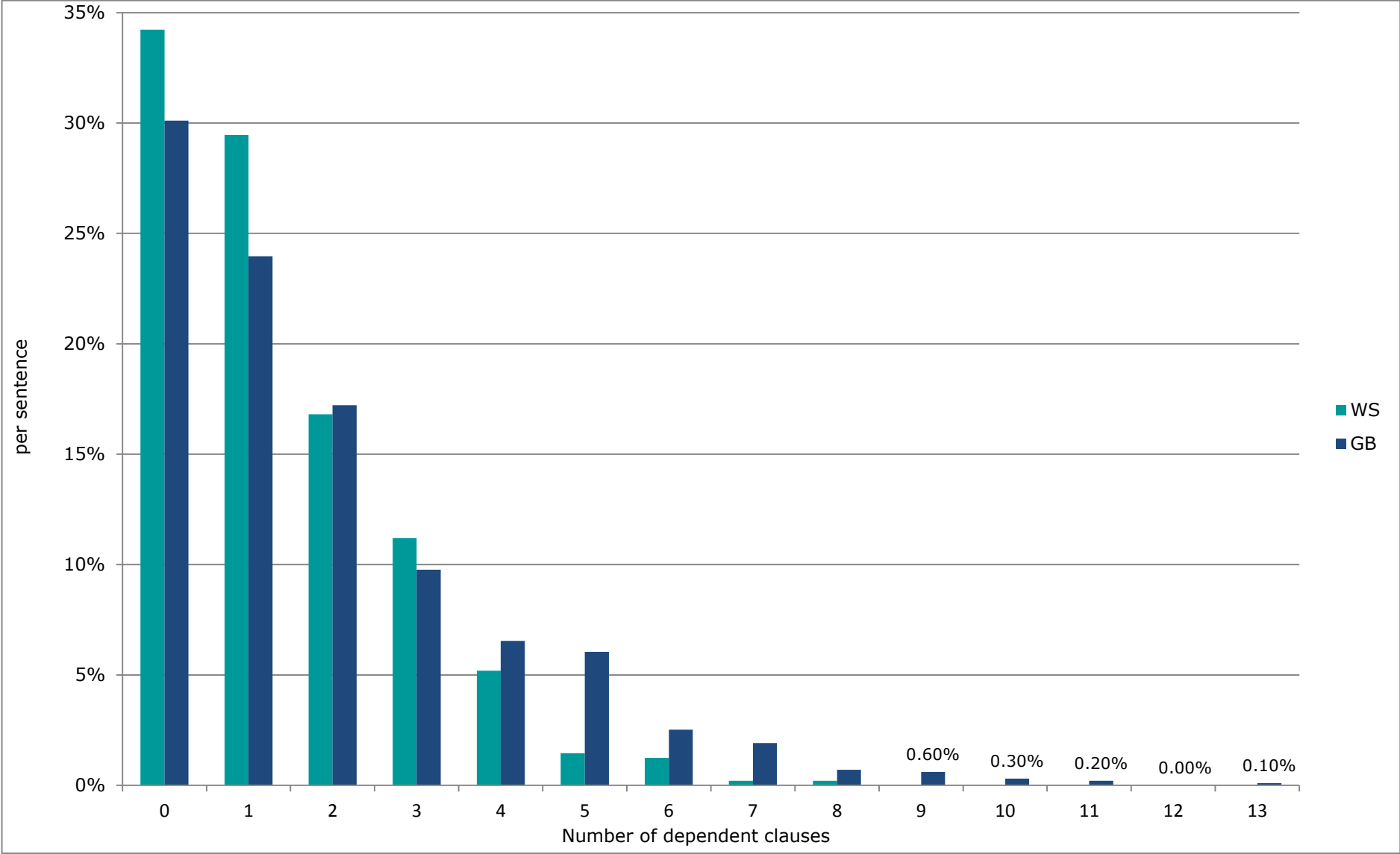
Figure 5-10 shows the distribution of dependent clauses across the *HJPC*, with the vertical axis indicating the percentage of sentences containing each number of dependent clauses for each part of the corpus. It has already been said that *GB* has many more dependent clauses than *WS* overall, but this graph provides a more detailed analysis of the contrast between the syntactic structure of *WS* and *GB* sentences. 34% of *WS* sentences have no dependent clauses and close to 30% more have one dependent clause. Over 91.69% of *WS* sentences have zero to three dependent clauses. 8.3% of *WS*'s sentences have four to eight dependent clauses, although only two sentences have more than six dependent clauses and none has more than eight. In *GB*, 30% of sentences have no dependent clauses and 23% have one dependent clause. 81.07% of sentences have zero to three dependent clauses. While these numbers are lower than those for *WS*, more than four fifths of *GB*'s sentences have the same clausal structure as the overwhelming majority of *WS*'s sentences. However, a great difference is seen in the sentences with large numbers of dependent clauses. 17.72% of sentences in *GB* have four to eight dependent clauses, a real contrast with *WS*'s 8.3%. 1.2% of *GB*'s sentences have nine to 13 dependent clauses. So, while most sentences in *GB* have a similar sentence structure to those of *WS*, there are sentences in *GB* with a much greater amount of dependency than anything which appears in *WS*. This very small number



of sentences in *GB* with exceptionally complex syntactic structures gives an impression of a text which is complex overall when most of the text is identical to *WS* in terms of dependency.

The sentences within the text with very high levels of dependency could be considered to be foregrounded. They have unusual properties which seem likely to slow a reader's progress and might draw their attention. Essentially foregrounding is a description of reader response and this project does not extend to research on whether the impact of the sentences' meaning is enhanced (or decreased) by their unusual form. However, it seems from the very large number of critics and general readers who point out the difficulty they perceive in James's late style overall that these sentences influence readers' judgement of the text as a whole. This is effectively almost the opposite of foregrounding. The exceptional sentences do not stand out as different from those around them but rather colour the perception of James's late prose. It is not clear why this should be the case.

Figure 5-10 Distribution of dependent clauses



### 5.3 Multiple main clauses

As described in section 4.4.2, I have added the code <apu> (additional parsing unit) to the *HJPC* to indicate where James has used more than one element which could stand as an independent item – almost always another sentence. This is illustrated in example 37).

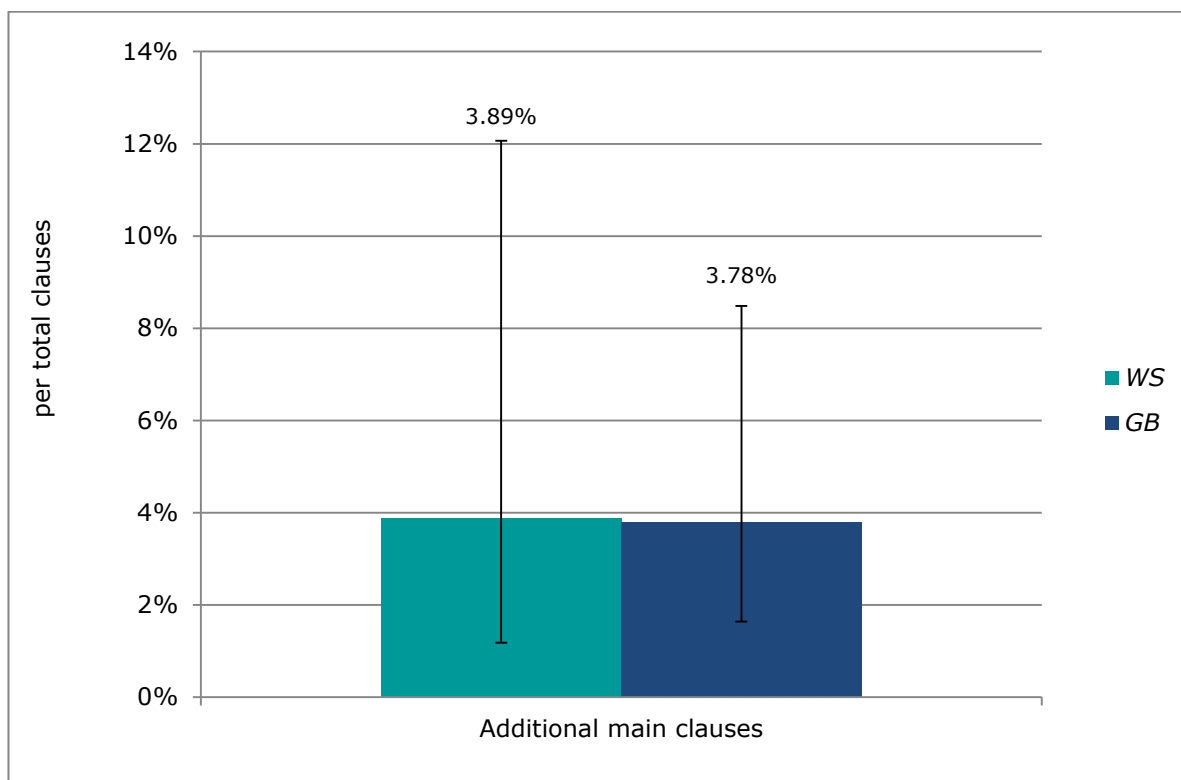
- 37) You asked acquaintances when you HAD your kith and kin -- <apu> you asked them over and above.

This construction does not have the embedded complexity described by Givón (2009), but neither does it have the clearly-signalled iterative construction of a compound sentence using coordinators. It seems to be intermediate in difficulty for the reader, requiring a recognition that another main clause is beginning, whether it is signalled in an unusual way by a dash, or more conventionally by a semi-colon or colon. It is therefore worth determining whether this is an additional complication in *GB* as compared to *WS*.

A simple search in ICECUP using the markup search facility chapter by chapter and compiling the results gives the total incidence of APU constructions for each part of the corpus. In order to normalise these results, they were then divided by the number of clauses in the sample and converted to a percentage result. Once the graph was generated, statistical significance was checked using the Wilson test and the margin of error indicated with error bars.

The result of the comparison is shown in Figure 5-11, which shows that APU is a relatively rare feature in both novels, with approximately 4% of clauses in each. Unlike all the other statistical analyses, the confidence intervals here are extremely wide and they overlap. Their overlapping indicates that no statistically significant difference has been demonstrated between *Washington Square* and *The Golden Bowl* with reference to APU's, as there is not enough data in the HJPC for such a conclusion. The wide confidence intervals show that it cannot be established that the data collected did not occur by chance. Contrary to my earlier hypothesis, APU's cannot be shown to be a distinguishing feature of the late style, a result which indicates the usefulness of checking subjective impressions with quantitative data. In fact, James's use of extra syntactically independent elements within an orthographical sentence is remarkably constant between *WS* and *GB*.

Figure 5-11 Multiple main clauses



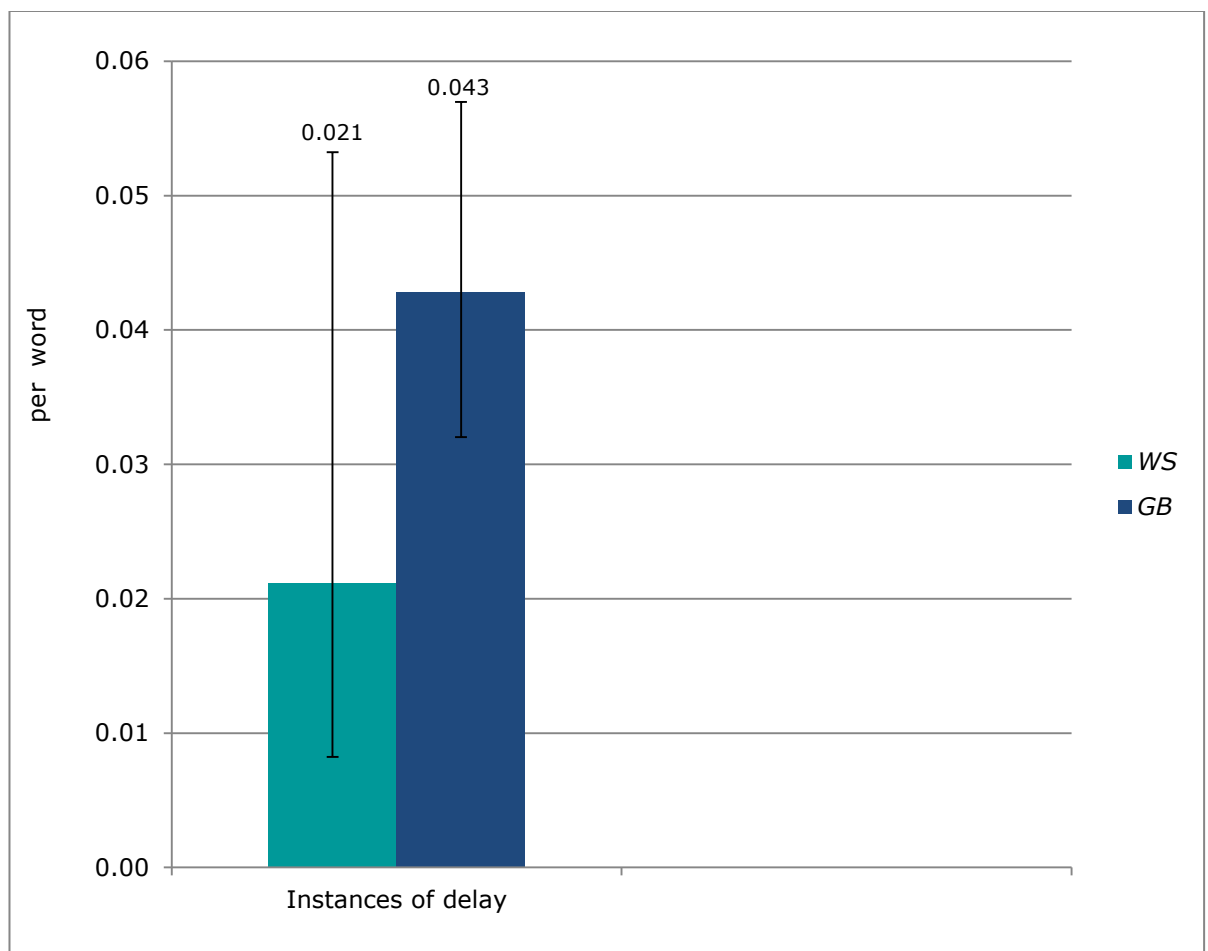
## 5.4 Delay

The phenomenon of delay, which was described and defined in section 4.4.3, provides a clear differential between the two samples. The *HJPC* was searched for instances of delay using ICECUP's markup search facility to find this code in the corpus. This provided data which was normalised using the total number of words in each part of the *HJPC*. The choice of total number of words as the basis for comparison avoids distortion of the data due to the variable length of sentences and clauses. The result is shown in Figure 5-12, which shows that *GB* has approximately twice the number of instances of delay per total number of words when compared to *WS*. However, the error bars are very large, particularly for *WS*; this is because the number of instances of delay (188) is very small compared to the total number of words in *WS* (8,888), resulting in overlapping confidence intervals.

The data for instances of delay show a marked difference between *WS* and *GB*. However, because the number of instances of delay is very low relative to the total number of words, the graph has very large and overlapping confidence intervals. Therefore there is a need to apply further statistical analysis to determine whether the observed difference between *WS* and *GB* is statistically significant. The Newcombe-Wilson test is used as an alternative to the Wilson test when the number of data points is low. The Newcombe-Wilson test recalculates confidence

intervals looking at the points of best and worst fit to ascertain whether these differences are statistically significant. In this case, there is a statistically significant difference between *WS* and *GB* in the instances of delay per word. I therefore conclude that the amount of delay is one of the defining differences between the early and late styles, and I suggest that the large quantity of delay in *The Golden Bowl* causes much of the unusual impact and difficulty of the late style, as discussed in section 4.4.3 with reference to Figure 4-9. The quantity of delay is related to the use of dependent clauses to some extent, but many instances of delay do not involve dependent clauses.

**Figure 5-12 Instances of delay per total words**



The importance of delay revealed through this corpus stylistic method echoes the comments of Ohmann (1964) in his article exploring literary style through the lens of generative grammar. Ohmann comments on James's short story 'The Bench of Desolation' (1909), finding that 'the style is idiosyncratic in the highest degree' (Ohmann, 1964, p. 436). Ohmann identifies embedding as the chief characteristic of James's style and suggests that it is a source of difficulty for the reader:

The embedded elements, in short, significantly outweigh the main sentences [sic] itself, and needless to say, the strain on attention and memory required to follow the progress of the main sentence over and around so many obstacles is considerable. ...It seems likely that much of James's later style can be laid to this syntactic device - a matter of *positioning* various constructions, rather than of favoring a few particular constructions (Ohmann, 1964, p. 437).

The data shown in Figure 5-12 also demonstrate that the late style is not a complete innovation on James's part. Rather it is a development of certain characteristics of his earlier style to suit his later purpose of focusing on his protagonists' inner lives through which the novel develops, rather than on the plot. James explains this approach in his *Preface to The Golden Bowl* in the New York Edition.

The Prince ... virtually sees and knows and makes out, virtually represents to himself everything that concerns us ... Having a consciousness highly susceptible of registration he thus makes us see the things that may most interest us reflected in it as in [a] clean glass. (James, 2000 [Original work published 1904], p. XVIII)

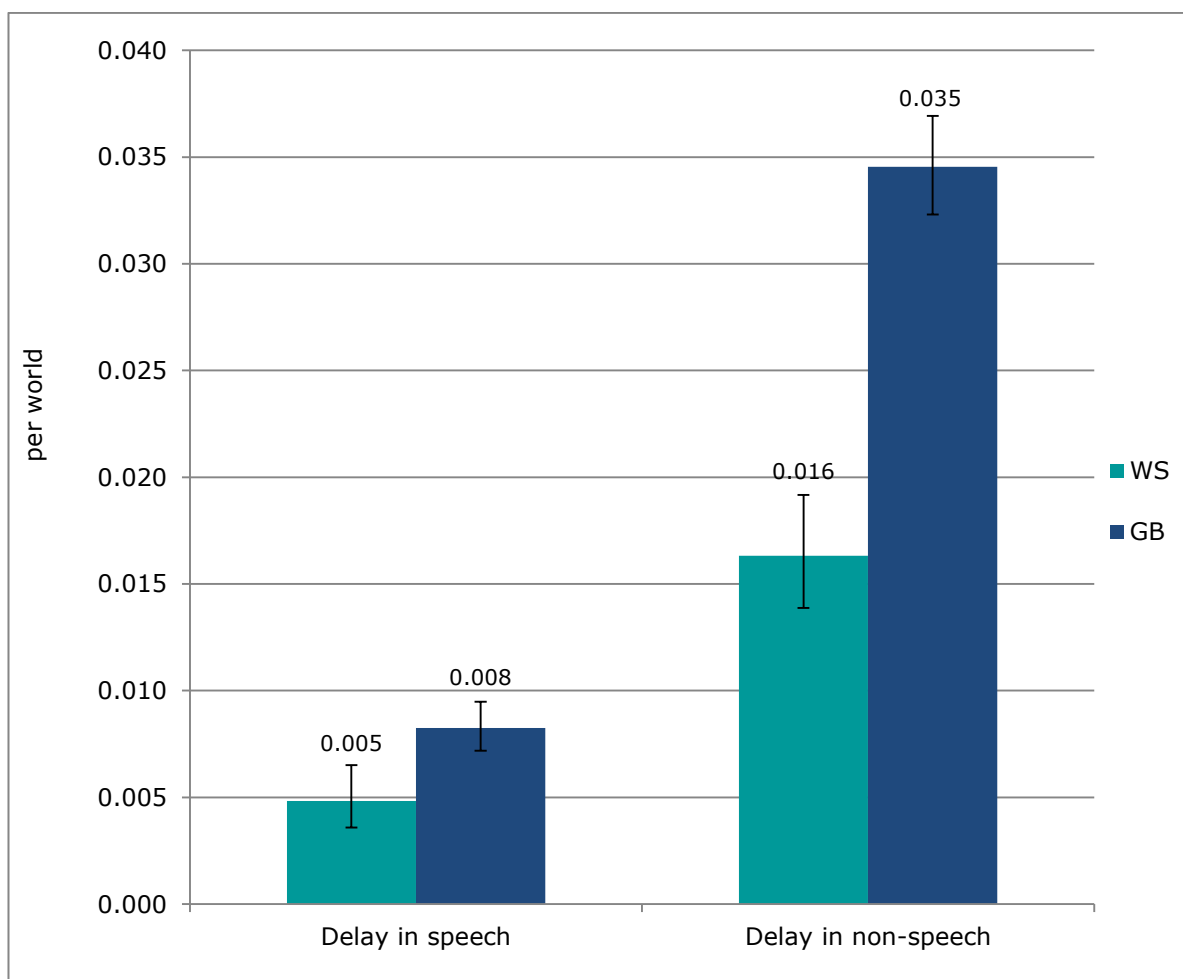
James goes on to explain that the same applies to the Princess in the second part of *The Golden Bowl*. I suggest that delay contributes to the effect James creates, illustrating the tendency for the mind to wander from thought to thought, rather than pursue a linear logic. So GB38:16, the exceptionally complex sentence quoted above as 36) and repeated below as 38), includes the parenthetical, marked in italics -- *however abysmal a face it had worn for her* --. The sentence is in free indirect thought and so superficially is from the viewpoint of the narrator. Nevertheless, this passage depicts Maggie's mental process and the parenthetical gives us her own feelings about the card game inserted into her memory and a description of how all four protagonists behaved and felt. The clause - *had found herself repeatedly doing so*; - appearing as a second verb phrase and complements within the first main clause, mimics in its form Maggie's compulsive return to this painful scene, not only in the actual words, but in the sentence form.

38) Maggie lived over again the minutes in question -- had found herself repeatedly doing so; to the degree that the whole evening hung together, to her aftersense, as a thing appointed by some occult power that had dealt with her, that had for instance animated the four with just the right restlessness too, had decreed and directed and exactly timed it in them, making their game of bridge -- *however abysmal a face it had worn for her* -- give way precisely to their common unavowed impulse to find out, to emulate Charlotte's impatience; a preoccupation, this latter, attached detectedly to the member of the party who was roaming in her queerness and was, for all their simulated blindness, not roaming unnoted. [my italics] (GB38:16)

### 5.4.1 Delay and the representation of speech

As shown in section 5.2.3, James alters the style of his prose considerably when depicting direct speech. Figure 5-13 shows the instances of delay as seen in Figure 5-12, but this time separated into those which occur in speech sentences and those in narrative sentences. In both parts of the *HJPC*, speech is marked by including fewer instances of delay. There is a greater differentiation between the two types of discourse in *GB* – the ratio of speech to non-speech instances of delay is 1:3.2 in *WS* and 1:4.38 in *GB*. This shows James’s greater effort in the later novel to represent natural conversation between his characters, which may be partly the result of his work in drama which he undertook in the 1890’s. In non-speech James uses 2.19 instances of delay in *GB* for each instance in *WS*, whereas in speech the ratio is only 1.6. This reinforces the finding that it is in non-speech sentences where James’s late style is mainly seen, both in terms of dependency and delay.

**Figure 5-13 Instances of delay in speech and non-speech**



## 5.4.2 Exceptional sentences

As with complexity, sentences with a very large amount of delay are likely to have a disproportionate effect on the reader. Table 5-4 shows the sentence in each chapter with the highest delay score. Note that this is not a count of instances of delay as in Figure 5-12 and Figure 5-13, but rather the table shows the count made of how many words are used to cause delay in the sentence overall, which may occur in one or more syntactic constructions. This is recorded at the beginning of each sentence in markup code, as ICECUP does not provide a counting facility for custom-made annotations. This makes it possible to scan a chapter picking out the exceptional sentences. The text of the sentences is shown below the table with the delay code included to clarify the calculations.

The contrast between the two novels is clear. The highest scoring sentence in *WS* has a delay score of 27, but this is quite exceptional with the other high scores being 13 or 15. The average of this collection of exceptional sentences is 24.2. This is far lower than any of the highest scoring sentences in the chapters from *GB*. Here the range is from 38 to 55 with an average of 47.5. Like the instances of delay figures, *GB* scores at approximately double the level of *WS*.

**Table 5-4 Exceptional sentences (2)**

Sentence number	Highest delay score
WS01:23	15
WS02:4	13
WS16:27	13
WS18:112	13
WS31:31	27
<b>Average</b>	<b>24.2</b>
GB01:55	55
GB20:28	38
GB22:2	47
GB38:75	50
<b>Average</b>	<b>47.5</b>

In these examples the words causing the delay are underlined. The code in grey after the delay indicates the number of words causing it and the function of the word which has been delayed. A key to the abbreviations for word functions



included in these sentences is provided immediately after them. The code <Dx> in grey at the beginning of each sentence consolidates the individual delay scores in the sentence, with x being the total delay score for that sentence. For further details, see section 4.4.3.

- 39) <D15> Of course his easy domestic situation saved him a good deal of drudgery, and his wife's affiliation to the "best people" brought him a good many of those patients whose symptoms are, if not more interesting in themselves than those of the lower orders, at least more consistently <delay> <15> <MVB> displayed.<sup>45</sup> (WS01:23)
- 40) <D13> She bloomed herself, indeed, and was a comely, comfortable, reasonable woman, and a favourite with her clever brother, who, in the matter of women, even when they were nearly related to him, <delay> <13> <VB> was a man of distinct preferences. (WS02:4)
- 41) <D13> Mrs. Penniman was silent a little, and her smile beneath the shadow of her capacious bonnet, on the edge of which her black veil was arranged curtain-wise, <delay> <13> <VB> fixed itself upon Morris's face with a still more tender brilliancy. (WS16:27)
- 42) <ds> <D13> "There is one thing you can tell Mr. Townsend, when you see him again," he said, <delay> <7> <OD> "that if you marry without my consent, <delay> <6> <SU> I don't leave you a farthing of money. (WS18:112)
- 43) <D27> The letter was beautifully written, and Catherine, who kept it for many years after this, <delay> <8> <VB> was able, when her sense of the bitterness of its meaning and the hollowness of its tone had grown less acute, <delay> <19> <AJPO> to admire its grace of expression.(WS31:31)
- 44) <ds> <D9> "I don't feel, my dear, if you really want to know, <delay> <8> <OD> that anything much can now either <delay> <1> <MVB> hurt me or help me. (GB01:55)
- 45) <D38> What any one "thought" of any one else -- above all of any one else WITH any one else -- was a matter incurring in these halls <delay> <3> <OD> so little awkward formulation that hovering Judgement, the spirit with the scales, <delay> <5> <VB> might perfectly <delay> <1> <AVB> have been imaged there as some rather snubbed and subdued but quite trained and tactful poor relation, of equal, of the properest, lineage, only of aspect a little dingy, doubtless from too limited a change of dress, <delay> <20> <FNPPPO> for whose tacit and abstemious presence, never betrayed by a rattle of her rusty machine, <delay> <9> <SU> a room in the attic and a plate at the side table were decently usual. (GB20:28)
- 46) <D47> Its general brightness was composed doubtless of many elements, but what shone out of it as if the whole place and time had been a great picture, from the hand of genius, <delay> <5> <FNPPPO> presented to him as a prime ornament for his collection and all varnished and framed to hang up --

<sup>45</sup> A version of this sentence with less delay could be: '...whose symptoms, if not more interesting in themselves than those of the lower orders, are displayed more consistently at least'.

what marked it especially for the highest appreciation <delay> <42> <VB> was his extraordinarily unchallenged, his absolutely appointed and enhanced possession of it. (GB22:2)

[Within the long delay in this sentence is a shorter one occasioned by the phrase *from the hand of genius*. *As if* is counted as one word in ICECUP, forming a two part subordinator.]

47) <D50> Charlotte 's one opportunity meanwhile <delay> <1> <NPPO> for the air of confidence she had formerly <delay> <1> <MVB> worn so well and that agreed so with her firm and charming type <delay> <16> <VB> was the presence of visitors never, as the season advanced, wholly <delay> <5> <VB> intermitted -- rather in fact, so constant, with all the people who turned up for luncheon and for tea and to see the house, now replete, now famous, <delay> <21> <AJPO> that Maggie grew to think again of this large element of "company" as of a kind of renewed water-supply for the tank in which, like a party of panting gold-fish, <delay> <6> <SU> they kept afloat. (GB38:75)

### Key to abbreviations in these sentences:

AJPO	adjective phrase postmodifier
AVB	auxiliary verb
FNPPPO	floating noun phrase postmodifier
MVB	main verb
OD	direct object
SU	subject
VB	verbal (verb phrase)

### 5.4.3 Distribution of delay scores

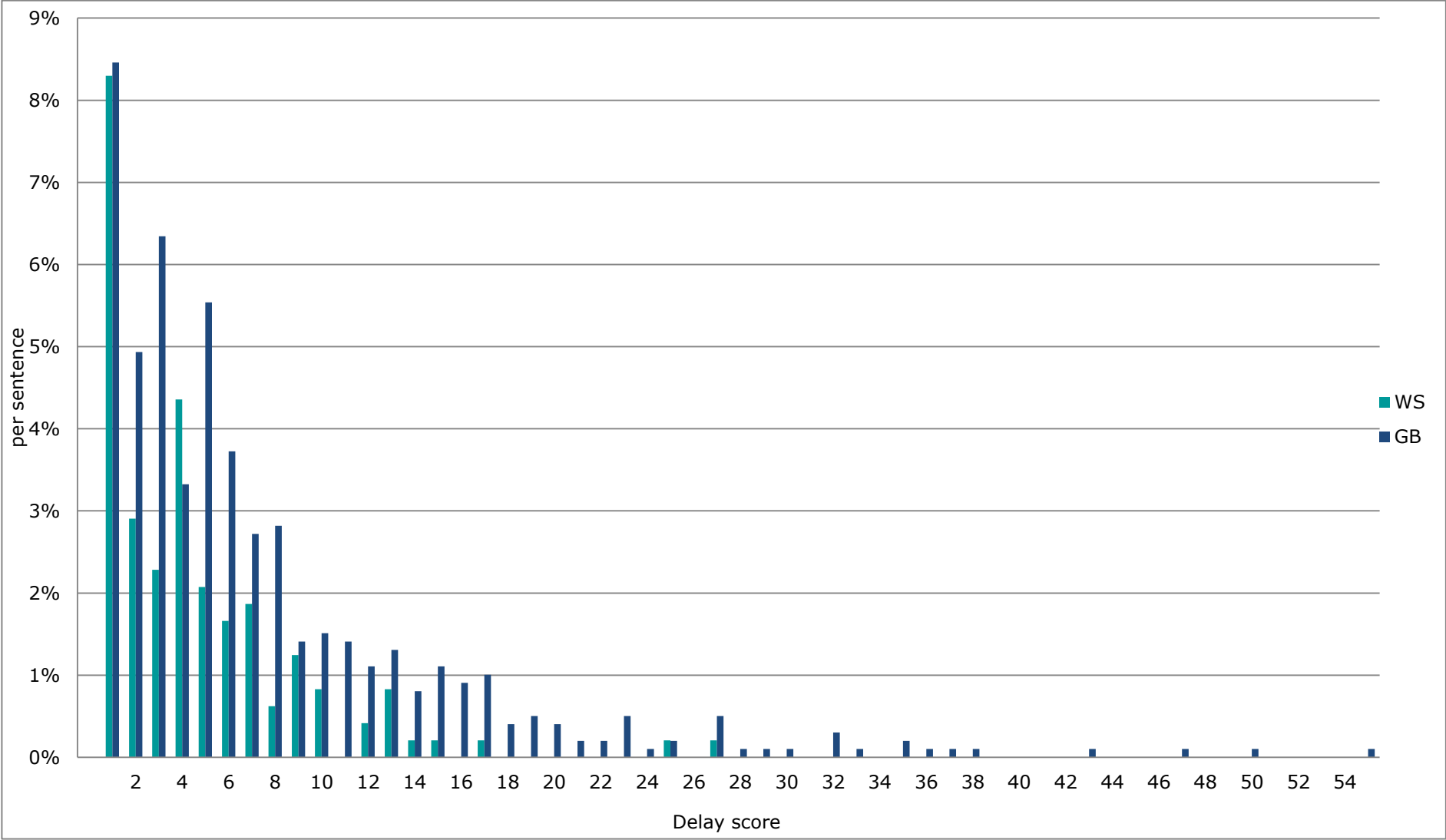
As with the sentences with high numbers of dependent clauses, it is important to emphasise that these sentences with exceptional delay scores do not reflect a general pattern. Rather they are unusual and stand out from the text. Figure 5-14 shows the distribution of delay scores in *WS* and *GB*. To compile these figures it was necessary to transfer them by hand from the *HJPC* in ICECUP into a table. The result shows the delay scores along the horizontal axis with the percentage of sentences with that score on the vertical axis. So, for instance, 4.35% of *WS* sentences have a delay score of 4 whereas only 3.32% of *GB* sentences have that score. The majority of sentences (71.48%) in *WS* have no delay at all. In contrast, in *GB*, only 46.93% of sentences have no delay. While Figure 5-14 confirms the greater amount of delay James uses in *GB*, the graph also makes it clear how isolated the high scoring sentences are. 85.06% of sentences in *WS* and 66.66% of sentences in *GB* have delay scores of zero to three. A difference emerges at

moderate levels of delay; 12.66% of *WS* sentences have a delay score of four to 10, whereas *GB* has 21.05% of sentences at this level. A real contrast is seen at the higher level of delay scores of 11 to 15. Only 1.66% of *WS* sentences reach this level whereas there are 5.74% of *GB* sentences of this type. Finally, only 3 sentences (0.62%) in *WS* delay scores of 16 and above but *GB* has 6.55% of sentences in this category. The effect of these unusual sentences with their fractured syntax is analogous to those with very high amounts of dependency;<sup>46</sup> the reader gains the impression of a text which is difficult to read because a feat of memory is required to follow the argument of the sentence. In fact, this occurs quite rarely but colours the overall impression of the novel.

---

<sup>46</sup> Delay score and amount of dependency are also related quantitatively, as some delay is delivered with the use of dependent clauses.

Figure 5-14 Distribution of delay scores



## 5.5 Conclusion

In this chapter, exploration of the *HJPC* has yielded a number of results indicating differences between the styles of *WS* and *GB*, and ways in which the two novels are similar. The statistical method that was applied, namely Wilson's score interval, makes it possible to conclude that the results are valid not only for the two relatively small samples which have been parsed and compiled into the corpus, but for the two books as a whole.

An initial examination of the two samples showed that the *GB* part of the *HJPC* is much larger than *WS*; they include the same number of chapters but *GB*'s chapters are far longer. The sentences in *GB* are approximately twice as long as those in *WS*, but the length of the clauses in the two parts of the corpus are similar. The length of *GB*'s sentences is explained by their having more, rather than longer, clauses; there are 2.3 clauses in *GB* for one clause in *WS*. Thus the first characteristics of James's late style identified are the length of the chapters and sentences, and the quantity of clauses. This finding suggested that it would be worthwhile to explore in what discourse type the most clauses were used, and what type of clause was most frequent.

While 'difficulty' is not synonymous with syntactic or clausal complexity, being a response from a reader, rather than an objective measurement, great syntactic complexity is likely to be a component of reader difficulty and therefore worth analysis in this context. Syntactic complexity at the clausal level was defined as the use of dependent clauses, which was contrasted to sentences containing coordinated clauses. Analysis showed that *WS* has approximately 2 dependent clauses for every coordinated clause whereas for *GB* the ratio was 3.6 to 1, indicating that an increase in the use of dependency, i.e. syntactic complexity, was associated with the later novel. This preliminary analysis also confirmed that coordinated clauses (analogous to Cross's concept of compounding, discussed in section 3.2.5) form a greater percentage of all clauses in *WS* than in *GB* (20.8% versus 15%). The ratio of main to coordinated clauses is higher in the early than the late novel; this comparison in *WS* is 2.32:1 whereas in *GB* it is 2.63:1.

However, the type of discourse which James used in each novel, and the extent to which he used it, was likely to be a factor in the number of dependent clauses the texts contain. Speech being generally believed to involve less dependency, the representation of speech might also have a reduced level of dependent clauses. For this reason, the relative quantities of speech representation in the two samples were important. On calculating these, it was shown that, while the two novels have similar amounts of sentences containing speech, there are

more speech clauses in *WS* than in *GB*. Therefore, an accurate comparison of James's early and late styles could only be confidently given if speech and non-speech sentences were shown separately. Then it was necessary to explore whether James did indeed use less dependency when representing direct speech and this proved to be the case. Once the two kinds of sentences were separated out, the true difference between the narrative sentences could be seen. In speech sentences *WS* and *GB* have almost the same ratio of main to dependent clauses. However in non-speech approximately 59% of clauses are dependent in *GB*, whereas in *WS* it is only 45%. There was also a markedly greater contrast in regard to the use of dependency between speech and non-speech sentences within *GB* than within *WS*, showing that in the late novel James differentiated the two types of discourse more than he had done before. This might be a consequence of his experience of writing dialogue in drama during his 'theatrical years'.

These results show that James's late style, as represented in *The Golden Bowl*, is more syntactically complex than the style of *Washington Square* but only in non-speech sentences. However, even within non-speech the majority of sentences have a similar clausal structure in both novels. The late style is characterised by a high level of syntactic complexity in a small number of sentences. This is more difficult for readers to process and has created some of the impression of the 'difficulty' which critics describe. However, in speech sentences no particular late style exists in relation to clause type.

A small number of sentences in *GB* have exceptional numbers of dependent clauses. There is a strong contrast here with *WS*, where the sentences with the greatest amount of dependency are much less extreme. 12% of sentences in *GB* have more than four dependent clauses (in *WS* the figure is 3%) and 1.2% have nine dependent clauses or more. These syntactically highly complex sentences are rare but many critics describe the style as homogeneously 'difficult'. I infer from this data that these exceptional sentences make a disproportionate impact on the reader, providing difficulties in interpretation and giving a sense that the text is more syntactically complex than it really is overall.

In contrast, the exploration of the phenomenon of sentences in which there is more than one syntactically independent element did not prove to be a distinguishing characteristic between the two novels. There is no statistically significant difference between them in this respect. This example illustrates the importance of the use of statistical data in the description of literary style. My initial subjective impression when parsing the corpus, and therefore dealing with each sentence in great detail, was that this was a dominant phenomenon in *GB* as compared to *WS*, but a search of the data shows otherwise.

Another phenomenon which was first identified subjectively, delay, did prove to differentiate the two samples very clearly. Taking the two novels as a whole, a very clear distinction can be seen between the instances of delay in *WS* and *GB*. *GB* has approximately double the number of instances of delay per word compared to *WS*. A more detailed analysis showed that this high level of delay occurs mainly in non-speech sentences. James uses 2.19 instances of delay in *GB* for each instance in *WS* in non-speech, whereas in speech the ratio is only 1:1.6. As with syntactic complexity, James's late style is found only in non-speech sentences. Where speech is depicted, *GB* resembles *WS* closely. Again James marks the representation of speech by simplifying his language and lessening the number of times he introduces delay. As with dependent clauses, there is a much clearer difference between the use of delay for speech and narrative in *GB* than there is in *WS*.

Some sentences have very large amounts of delay, but they make up a small percentage of the whole. The great majority of sentences in *GB* are similar in terms of delay to those in *WS*. *GB* has substantially more sentences than *WS* with a moderate level of delay and only tiny numbers of *WS* sentences have delay scores of 11 or above. James's use of parenthesis is widely commented on by critics such as Watt, Chatman and Smit (see section 3.2.6) as though this were a homogeneous feature of the late style. Again, I infer that these exceptional sentences make a disproportionate impact on the reader, giving the impression that much of *The Golden Bowl* has large amounts of parenthesis.

Altogether the data which has been retrieved from *HJPC* points to a style in *The Golden Bowl* which has more syntactic complexity and delay compared to *Washington Square*, but only in the sentences which do not include direct speech. Even if the late style is taken to apply only to non-speech, most of *The Golden Bowl's* sentences are no more complex than those of *Washington Square*. If the two novels used in this project accurately represent James's early and late styles, James's late style can be characterised as complex in a minority of sentences. These complex sentences are formally externally foregrounded, that is, in Leech's (2008) terms,<sup>47</sup> they are a departure from the accepted norm for literary writing, as many critics protested from James's time onwards. The same sentences are also internally deviant, in that they are unlike the majority of sentences in *GB*. However, it is unclear whether they are functionally foregrounded, whether their complexity makes a particular contribution within the novel. This can only be determined by a more detailed qualitative analysis, of the type carried out in Chapter 6.

---

<sup>47</sup> See section 1.2.2

It should also be noted that the analysis in this chapter is not a complete characterisation of the syntax of *Washington Square* and *The Golden Bowl*. Many other syntactic elements could be analysed, and Hoover (Hoover, 2007) showed that James's lexis changed chronologically and differentiated three distinct phases of writing. Other elements, such as vague pronominal reference and the use of protracted metaphors, add to the difficulties readers encounter in reading James's late novels.

A different approach to the data in the *HJPC* will be discussed in Chapter 6, where the central chapter of each book will be explored in more detail.



# Chapter 6 Corpus Close Reading

---

## 6.1 Introduction

Chapter 5 presented the results of a corpus-aided analysis of *Washington Square* and *The Golden Bowl*, looking at each novel as a whole. The focus was on aspects of sentence structure and syntax, taking a statistical approach, and the research question addressed was the contrast between James's early and late styles. Chapter 6 presents a very different use of corpus stylistics, utilising quantitative data as an aid to a qualitative analysis. It offers a response to Bailey's criticism of corpus stylistics as mere 'data retrieval' (Bailey, 1989, p. 5), moving away from a linear model to one in which corpus data foregrounds key moments in the text which are thus highlighted for critical analysis.<sup>48</sup> This chapter also differs from Chapter 5 in that it explores internal deviation. Various *profiles* are compiled which display patterns across a chapter. These profiles are then used to identify *salient sentences* by measuring sentence length, coordination and complexity, and these prove to be helpful in elucidating both plot and character. 'Salient' here is used to mean sentences which are identified as among the most exceptional in their respective chapter by these criteria. Salient sentences are foregrounded because they are internally deviant, standing out from the majority of sentences in the chapter, and also because this deviation seems to be motivated, revealing aspects of plot or character. The exact cut off point for classification as salient is given in the relevant sections below.

A critical analysis, even using corpus stylistics, is, to some extent, a subjective matter, but here remains rigorous insofar as it links literary interpretation closely to a linguistic analysis of the text. The methodology of this chapter attempts to be more objective than that described by Halliday (1971) and summarised in section 1.2.2. Halliday identified prominent linguistic features and themes intuitively before quantifying them. Here, prominent features are identified by quantitative analysis and then their stylistic significance is considered. So, while the salient sentences are identified objectively, the interpretation is my own.

For such an analysis a smaller unit of text must be used to avoid the process of analysis becoming impossibly lengthy, analogous to Watt's analysis of only the first paragraph of *The Ambassadors* – see section 3.2.6. The text used here is one chapter from each novel. The chapters chosen are the numerically central chapters

---

<sup>48</sup> Bailey's critique of corpus stylistics was presented in section 2.4

of each book – Chapter 18 of *Washington Square* and Chapter 22 of *The Golden Bowl*, hereinafter WS18 and GB22.<sup>49</sup>

## 6.2 The two chapters

*Washington Square* was first published in six monthly parts, first in London's *Cornhill Magazine* from June to November 1880 and then in *Harper's Monthly* in New York from July to December of the same year. The version I have used to compile the WS section of the HJPC (*Henry James Parsed Corpus*) is an e-text made from the book published by Macmillan in London in 1881, which combined *Washington Square* with *The Pension Beaurepas* and *A Bundle of Letters* in two volumes. The text of Chapter 18 in this e-text, and from the Macmillan edition, is identical to the *Cornhill Magazine* version except for some punctuation changes shown as footnotes in Appendix 1. The effect of these is to split two relatively long sentences in the *Cornhill Magazine* into four shorter ones in the Macmillan text and thus the HJPC. This affects slightly the results shown in this chapter. More details are given in the sentence analyses in section 6.5.2. Chapter XVIII was the last chapter of the August section of the serialisation in *Cornhill Magazine* and also ended the September instalment in the American serial. It is the central chapter of the novel, which has 35 chapters in all. Its position as the end of an instalment may be significant in the plot and/or style of the chapter; its dramatic, forward-looking ending with Doctor Sloper determining 'to see it out' in the final sentence of WS18 encourages the reader to buy the next edition of the magazine.

Chapter 18 of *Washington Square* is not only the central chapter of the novel numerically but also represents a turning point in the development of the plot.<sup>50</sup> In this chapter Catherine confronts her father with her feelings about Morris Townsend, in her characteristically timid and submissive way, attempting to obey her father while not immediately abandoning her hopes of marriage. Her father accuses Townsend of being mercenary and makes it clear that he will never leave an inheritance to Catherine if she marries him. Faced with the horrifying idea of looking forward to her father's death, Catherine promises that if she does not marry Townsend in her father's lifetime, she will never do so after his death. This declaration is the turning point of her life; the novel ends with her still holding to her promise. The chapter is shocking in its cruelty, which is no less striking for being expressed in polite terms. Dr Sloper's feelings are nuanced; he cares about

---

<sup>49</sup> The texts of the two chapters, with the sentences numbered, are included at the end of this dissertation as Appendices 1 and 2.

<sup>50</sup> For a more complete plot summary, see section 3.3.1.

Catherine and wishes to protect her but is also amused by her attempt to be resolute. Like Catherine, he stays with the stance with which the chapter closes, matching his daughter's stubborn determination, for the rest of his life. Thus this central chapter crystallises the situation around which the rest of the plot will turn.

Chapter 22 of *The Golden Bowl* also represents a significant turning point in the relationship of the Prince and Charlotte, which then inevitably affects the other two main protagonists, Maggie and her father. In this chapter Charlotte and the Prince finally consummate their adulterous relationship when they are left behind at the end of a country house weekend.<sup>51</sup> The chapter is set at Matcham, where Charlotte and the Prince are guests while their respective spouses have elected to stay in London. It opens with the Prince strolling on the terrace in front of the house, ruminating on his relationships with women. The majority of the guests have left after the house party but Charlotte has been asked to stay on by their hostess, Lady Castledean, and the Prince has been asked to accompany her both there and on her journey back to London. Lady Castledean wants them with her so that she can also keep Mr Blint, her lover, at the house with apparent respectability. This part of the chapter is represented as the Prince's thought process. He reflects that he does not feel uncomfortable as a mere chaperone when other men, including Lord Castledean, have important matters to engage in. The reader understands that this internal monologue is not reliable as the Prince then describes himself as 'an outsider' who can 'among all these so often inferior people' be 'practically held cheap and made light of' (James, 2000 [Original work published 1904], p. 204). Charlotte appears at one of the windows overlooking the terrace, wearing a hat and jacket. For the Prince, this implies that she is ready 'to take with him some larger step' (James, 2000 [Original work published 1904], p. 205). He is also delighted that they simultaneously have the same thoughts and impulses; they are "meant for each other" (James, 2000 [Original work published 1904], p. 206). There is a brief conversation and Charlotte comes to join him. They admire the view of Gloucester and they discuss the 'gilded crystal bowl' (James, 2000 [Original work published 1904], p. 207) which they saw together before the Prince's marriage. Charlotte makes it clear that she is ready to go to nearby Gloucester with him, that she has already told her maid to go ahead of her to London, and that Lady Castledean doesn't want them to stay any longer. She even knows which train they would need to get from Gloucester to London in time to dine together with their spouses at her home, and which inn they should go to for lunch. The implication

---

<sup>51</sup> For a more complete plot summary, see section 3.3.2.

that they might find time to consummate their affair in Gloucester is not stated, but it is implied in their last exchange, when Charlotte says:

“These days, yesterday, last night, this morning, I’ve wanted everything.”

and the Prince’s responds:

Well, it was all right. “You shall have everything.” (James, 2000 [Original work published 1904], p. 210)

Whether or not this is the day on which their affair is first consummated, undoubtedly it is this decision which takes the Prince and Charlotte out of social convention and into the intrigue of an affair. Though there may be doubts about their staying at Matcham, they are chaperoned by Lady Castledean and Mr Blink. Once they leave alone for Gloucester, they are acting secretly. This is a crucial development in the novel and the lives of all the characters, signalled by the return to the topic of the golden bowl, which will be the key to Maggie’s realisation of the truth.

Thus, the numerically central chapter of each novel represents a significant turning point in the plot, although not the climactic point which comes later in each case: in *Washington Square* when Doctor Sloper bursts out with violent emotion against Catherine in a dramatic mountain setting, and in *The Golden Bowl* when Maggie confronts the Prince with her knowledge of his affair. The identification of these numerically central chapters as turning points could be considered to be circular. In fact, the chapters were chosen without any expectation of them having any particular role in the plot, but a subconscious bias cannot be entirely ruled out. This is an example of where personal critical judgement cannot be avoided, as discussed in the introduction to this chapter. A different reader might see the plot development of these two novels differently; it is my judgement that numerical and plot centrality coincide.

### 6.3 Chapter comparison

Table 6-1 shows some basic statistics for the two chapters to be analysed in detail. Although there are differences between the two parts of *HJPC* (*WS* and *GB*), *GB22* is comparatively similar to *WS18*; the ratio of words *GB22:WS18* is a little over 3:2 and sentence and clause ratios are even closer.

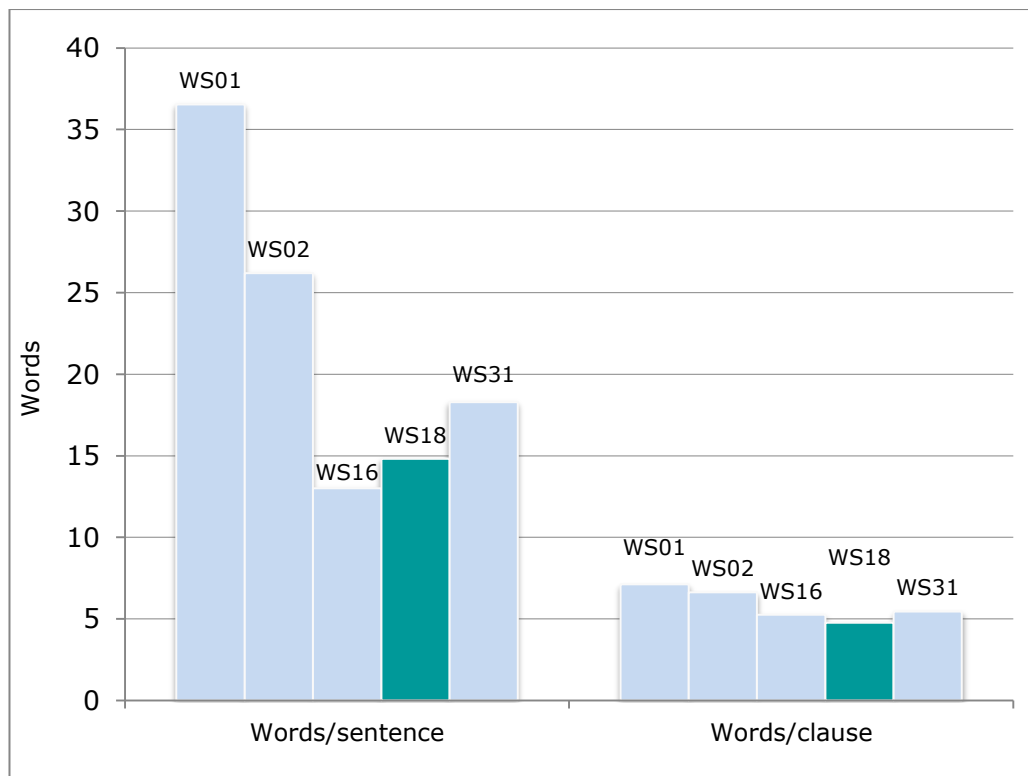
**Table 6-1 Comparative statistics**

	Words	Sentences	Clauses
--	-------	-----------	---------

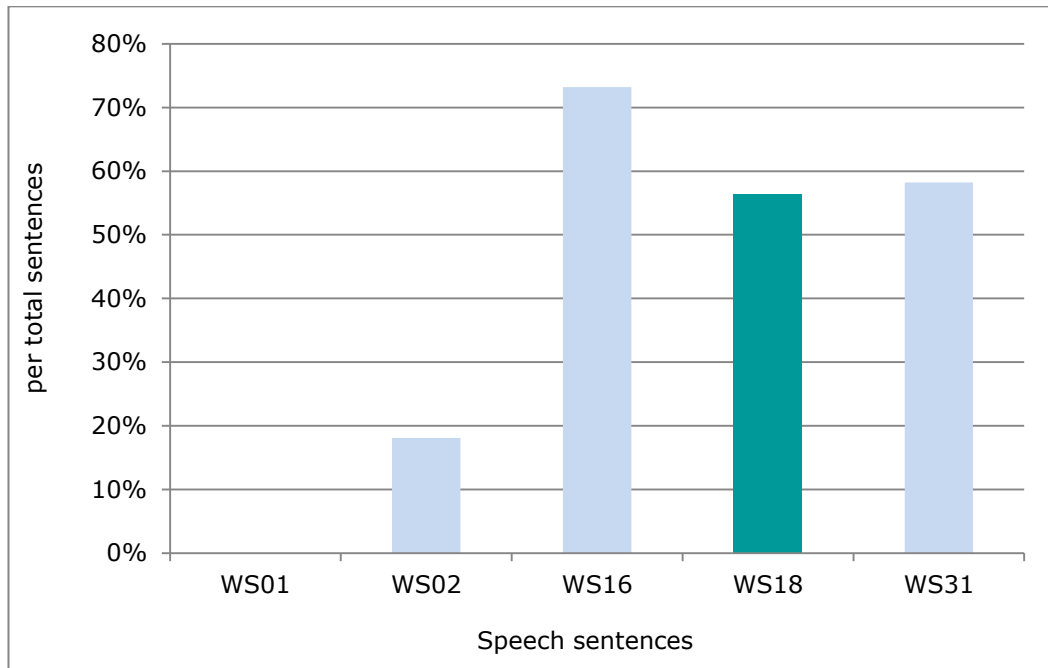
<b>Washington Square Chapter 18</b>	2073	140	435
<b>The Golden Bowl Chapter 22</b>	3770	165	552

Figure 6-1 shows WS18, highlighted in turquoise, in the context of the other chapters included in *WS*. While the number of words per sentence in WS18 is broadly in the middle of the range of values, it has the fewest words per clause of all the chapters analysed. This is not a reflection of a greater amount of speech in this chapter compared to the others, as shown by Figure 6-2. While WS01 has no speech at all, WS18 has no more than WS31 and considerably less than WS16. No reason has been identified for the relatively short clauses of WS18; it seems simply to be a stylistic variation.

**Figure 6-1 WS18 in context**

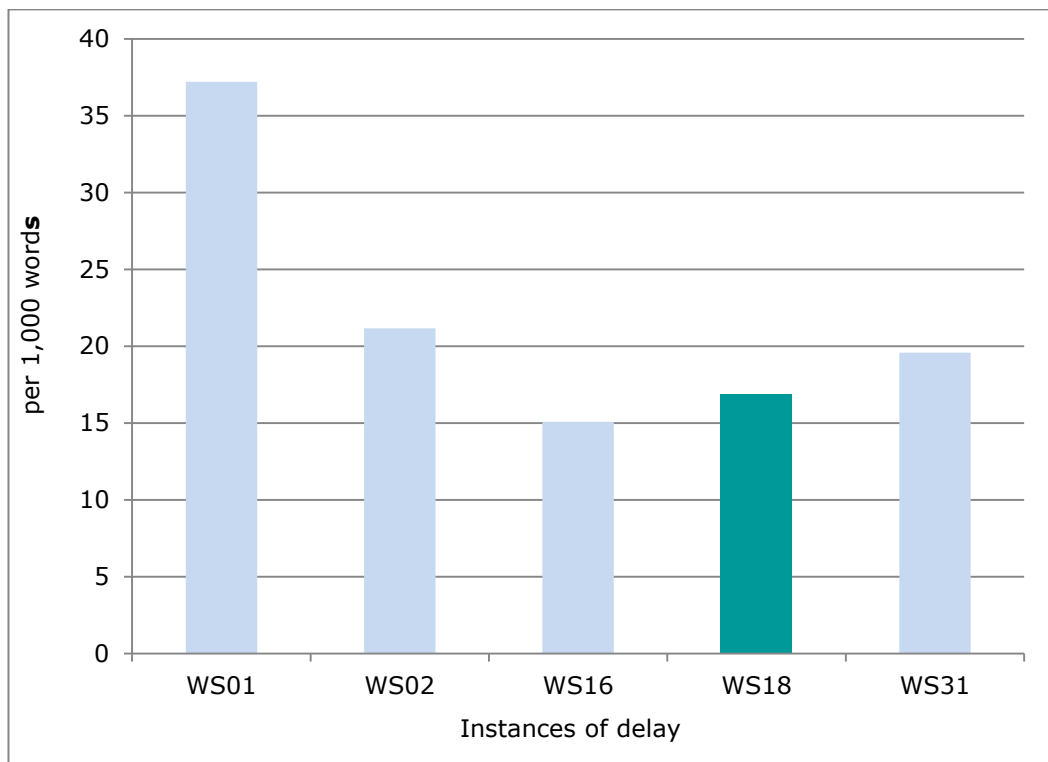


**Figure 6-2 Speech sentences in WS**

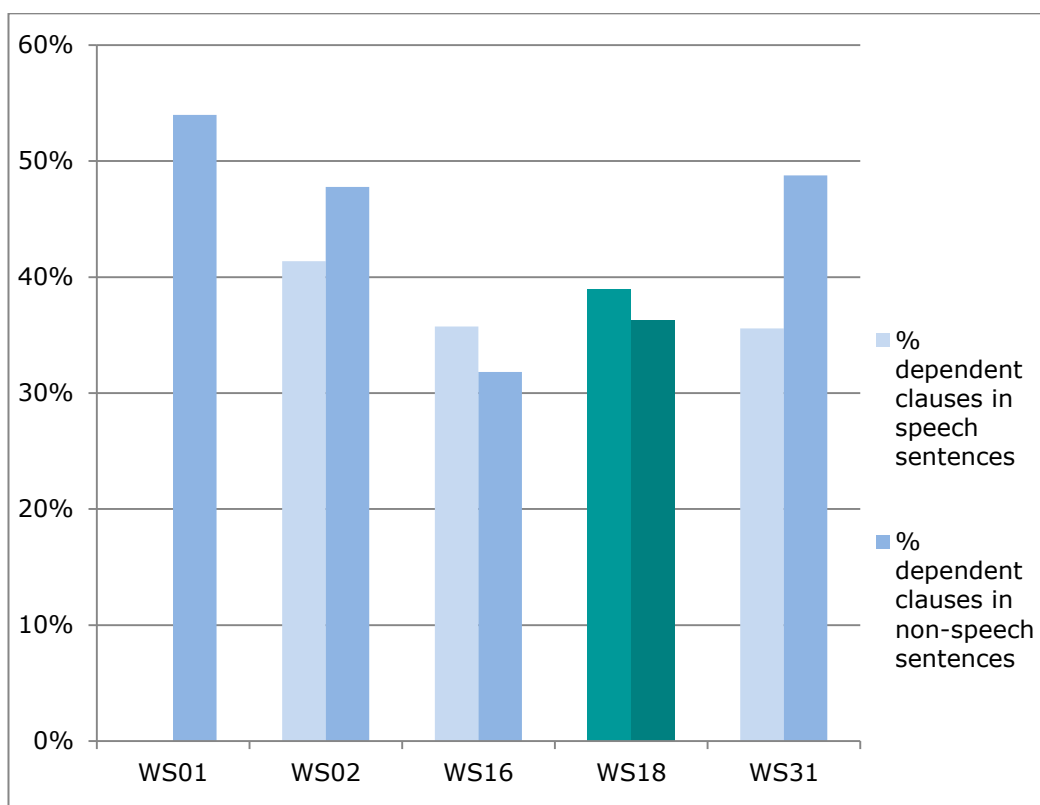


WS18 is unexceptional as compared to the other *WS* chapters in terms of delay, as shown in Figure 6-3 and in terms of the percentage of dependent clauses in speech and non-speech, as shown in Figure 6-4.

**Figure 6-3 Instances of delay per 1,000 words**



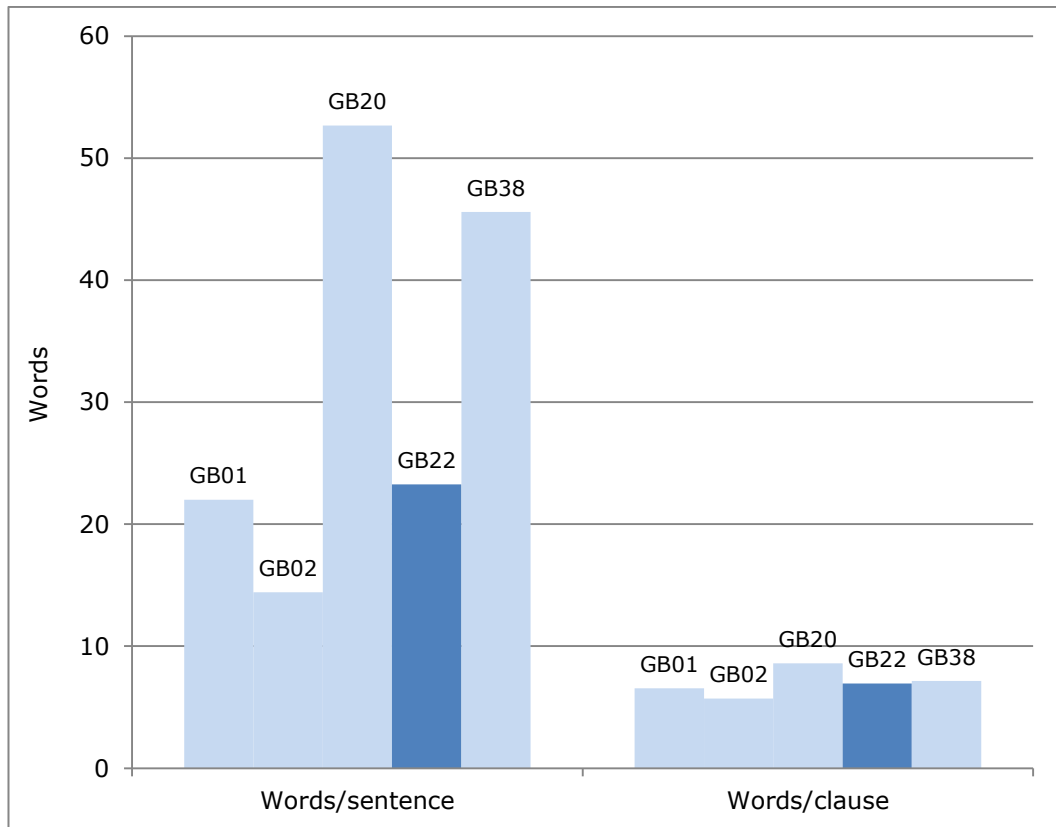
**Figure 6-4 Dependent clauses in speech and non-speech**



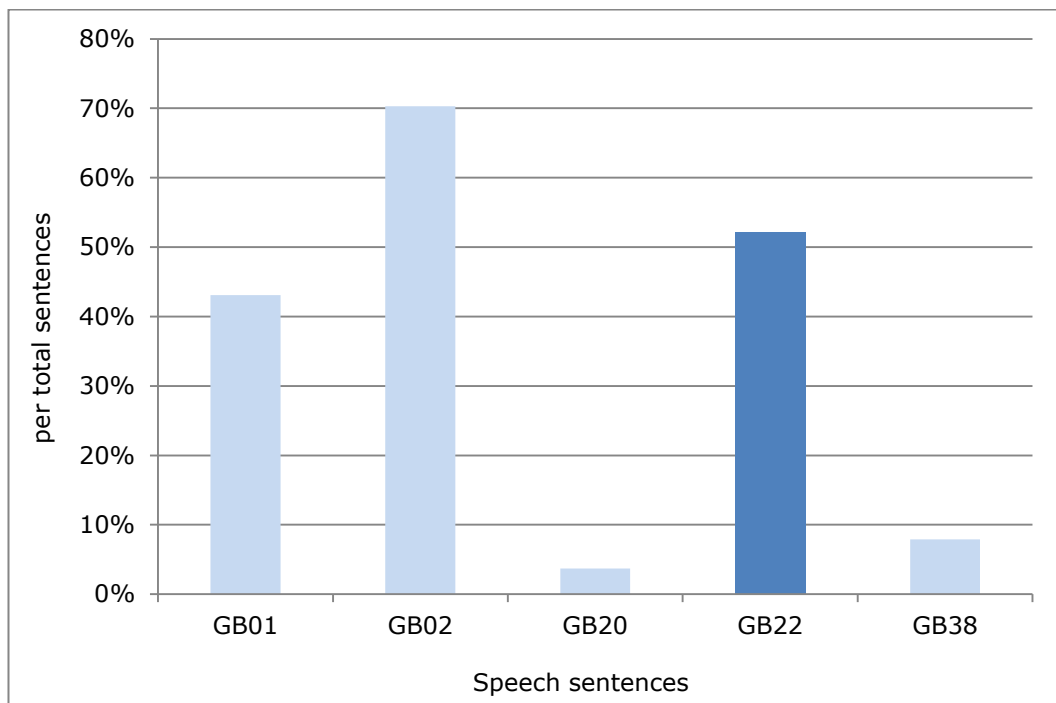
These four graphs demonstrate that WS18 is broadly comparable along the parameters measured with the other chapters of *WS*, so that its analysis is likely to have relevance for *WS* as a whole.

The following graphs explore the same comparisons for GB22. Figure 6-5 shows that GB22 has sentences in the middle to lower range of length when compared with the other chapters of *GB*; the clause length is relatively low but in that respect all the chapters are quite homogeneous, with words per clause only ranging from 5.71 to 8.59. However, GB22 is notable for the amount of dialogue it contains: 52.12% of all sentences in the chapter contain direct speech, only exceeded by GB02 with 70.31%. The amount of dialogue in each chapter is very variable, as is illustrated in Figure 6-6.

**Figure 6-5 GB22 in context**



**Figure 6-6 Speech sentences in GB**

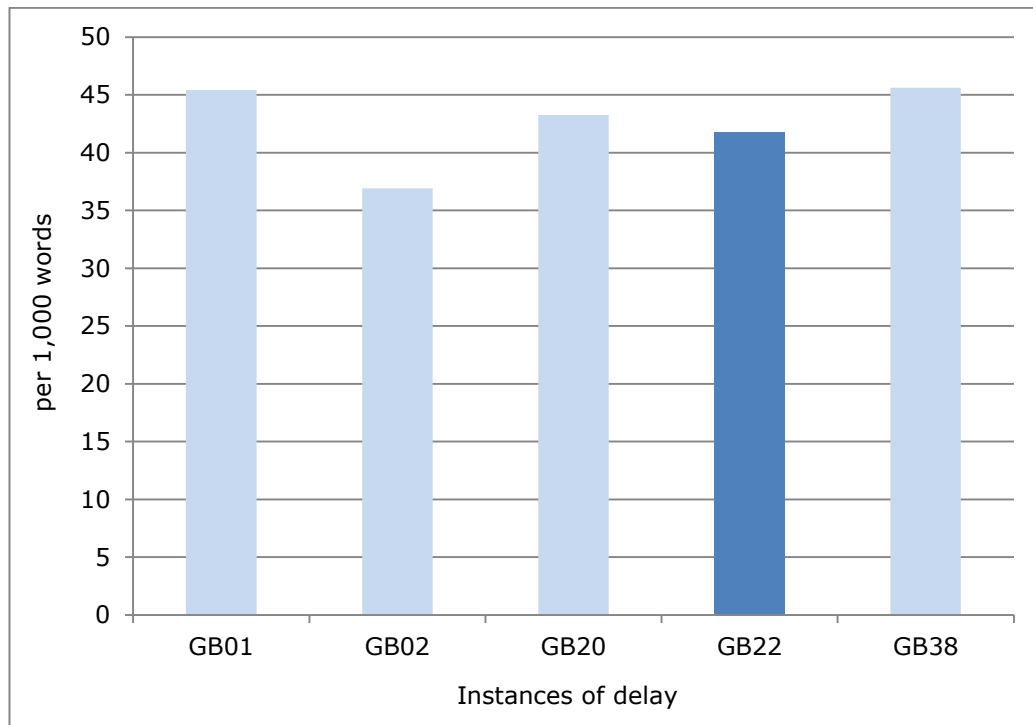


GB is particularly homogenous in terms of delay (see Figure 6-7). Although GB22 represents the second lowest level of delay at 41.71 instances per 1,000 words, the



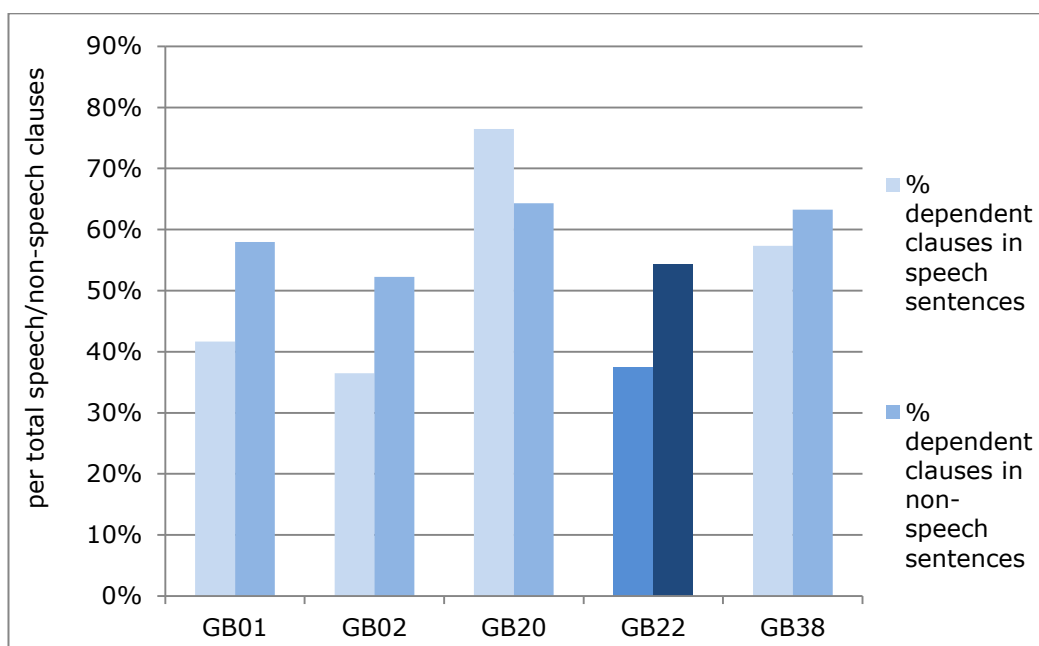
highest level (GB38) is only 45.60 instances per 1,000 words. James is extremely consistent in his use of this stylistic device.

**Figure 6-7 Instances of delay per 1,000 words**



There is a little more variation in his use of dependent clauses, measured as a percentage of all clauses. Speech and non-speech clauses are separated in Figure 6-8, where the relatively low level of dependency in the speech sentences of GB22 is clear. GB20 shows a different pattern to the other chapters of *GB* in the ratio of speech and non-speech dependent clauses. However, there are only two speech sentences in GB20, which is also a very short chapter. With such a tiny sample, an anomalous pattern of dependency is not surprising.

**Figure 6-8 Dependent clauses in speech and non-speech**



The two chapters discussed in detail in this analysis are thus broadly similar to the other chapters in *HJPC*.

## 6.4 Sentence comparison

This close textual analysis involves a different corpus stylistic technique to that used for the main study. Where whole texts were examined previously, now it is sentences which are compared. To undertake this task, a new methodology is required to characterise the sentences of *WS18* and *GB22* quantitatively. The simplest measure used is the number of words in the sentence, but coordination and complexity are also used to compile data.

For coordination, a *coordination score* is calculated for each sentence. This uses the total number of coordinated items in the sentence. For example:

- 48) (a) Her father sat looking at her, and (b) she was afraid he was going to break out into wrath; his eyes were so (c) fine and (d) cold. (*WS18:38*)

This sentence contains two coordinated clauses, (a) and (b), and two coordinated adjective phrases, marked (c) and (d). Counting these gives the sentence a coordination score of 4.

Similarly, a *complexity score* for each sentence was devised which includes a number of measures as discussed in Szmrecsanyi (2004) (see section 5.2.1. The

goal is to take into account a wide range of syntactic features which might add to a reader's difficulty. The factors included are:

- the number of dependent clauses in the sentence, as shown in red in example 49).
- 49) His assent could only be tacit, for he had never been dazzled by his sister's intellectual lustre. (WS02:16)
- the number of additional parsing units (APU) in the sentence. An APU is defined as a clause which could grammatically stand alone but which is joined with one or more similar clauses by James into a single sentence. For further details see section 4.4.2. In calculating the complexity score, the first parsing unit is not counted, as illustrated in example 50). The markup code <apu> is inserted into the corpus as each sentence is parsed and is here included in red.
- 50) "It doesn't matter," he said; <apu> "it's only for three or four years. (WS05:16)
- the number of 'detached function' clauses in the sentence, such as those used for a comment clause in the middle of speech, as explained in section 4.4.3 and illustrated in red in example 51).
- 51) "Well," rejoined Morris, "it is a fact that I wish to marry his daughter. (WS16:25)
- the delay score, being the total number of words causing delay in the sentence and indicated as <Dx> in the code at the beginning of the sentence, as shown in example 52). For more information on delay, see section 4.4.3. Here the main verb (MVB) *be* is delayed from the auxiliary verb *would* by the prepositional phrase *in that case*.
- 52) <D3> My value would in that case <delay> <3> <MVB> be estimated." (GB01:143)
- the number of paratactic clauses. Paratactic clauses are those in which there is direct speech, when the main clause of the sentence is the speech report, such as *he said*. This is shown in example 53) in red.
- 53) "Do you think it would be good for YOU?" Maggie Verver had smilingly asked. (GB01:53)
- the number of passive verb phrases

- the number of special marked clausal features in the sentence. Such features are not deviant in the sense of standing out as unusually complex. However, they represent an increase in complexity from the canonical clause defined in section 4.4.3 and so can each be counted in an assessment of the complexity of a clause.

These are:

- preposed items such as subject, direct object, indirect object, etc. Example 54) shows a preposed direct object in red.

54) I am not afraid of **what** your aunt will say when I go.” (WS05:98)

- extraposed subject or direct object (using the terminology of the ICECUP grammar). In example 55), anticipatory *it* (in red) is the provisional subject and the clause starting with *that*, also shown in red, is the notional subject. In the grammar used in ICECUP this is marked as extraposed, although other grammarians argue that the clause is not extraposed as it cannot be placed in the clause-initial position.<sup>52</sup>

55) **It** seemed to Catherine that if she were his sister she would disprove this axiom. (WS05:80)

- existential sentence, beginning with non-deictic *there is/are*, as in example 56).

56) **There are** plenty of sham ones about. (GB01:25)

- cleft sentence, e.g.

57) **it** had to be as covertly that he let his appreciation expand. (Part of GB22:115)

- pushdown, where an element of one clause is part of the syntactic structure of the following clause. In example 58) *what* is the direct object of *tell*, which is ditransitive.

58) “**What** did you say you would **tell** me?” she asked. (WS05:101)

- reduced structures, such as the direct object subordinate clause containing only *to* in example 59) which has no main verb.

---

<sup>52</sup> Throughout this dissertation, the grammar followed is that of ICE-GB which is, broadly speaking, that of Quirk et al. (1985). This is to facilitate consistency in the analysis.

59) "I should like to — but I am afraid I can't," Catherine answered.

The complexity score for each sentence is calculated by simply adding up the number of items in the above list with the exception of the delay score, where the score itself is added to the total. The calculation of a complexity score is illustrated by example 60), which has a complexity score of 22. The items which contribute to this total are marked in red. Coordination is marked in blue and the coordination score of this sentence is 8.

60) Castledean had gone up to London; [**<apu>**] the place was all her own; [**<apu>**] [**coordinated main clause**] she had had a fancy for a quiet morning with Mr. Blint, a sleek civil accomplished young man -- distinctly younger than her ladyship – [**delay of 5 between man and who**] [**dependent clause**] who [**coordinated verb**] played and [**coordinated verb**] sang delightfully ([**detached function clause**] [**coordinated dependent clause**] played even "bridge" and [**coordinated dependent clause**] sang [**coordinated noun phrase**] the English-comic as well as [**coordinated noun phrase**] the French-tragic), and [**coordinated main clause**] the presence – [**dependent clause**] which really meant the absence -- [**delay of 5 words between presence and of**] of a couple of other friends, [**dependent clause**] if they were [**passive**] happily chosen, [**delay of 5 words between friends and would**] would make everything all right. (GB22:13)<sup>53</sup>

Complexity score = 22

Coordination score = 8

15 Delay score <D15>

2 x main clauses

2 x APU's

2 x verbs

3 x dependent clauses

2 x noun phrases

1 x detached function clause

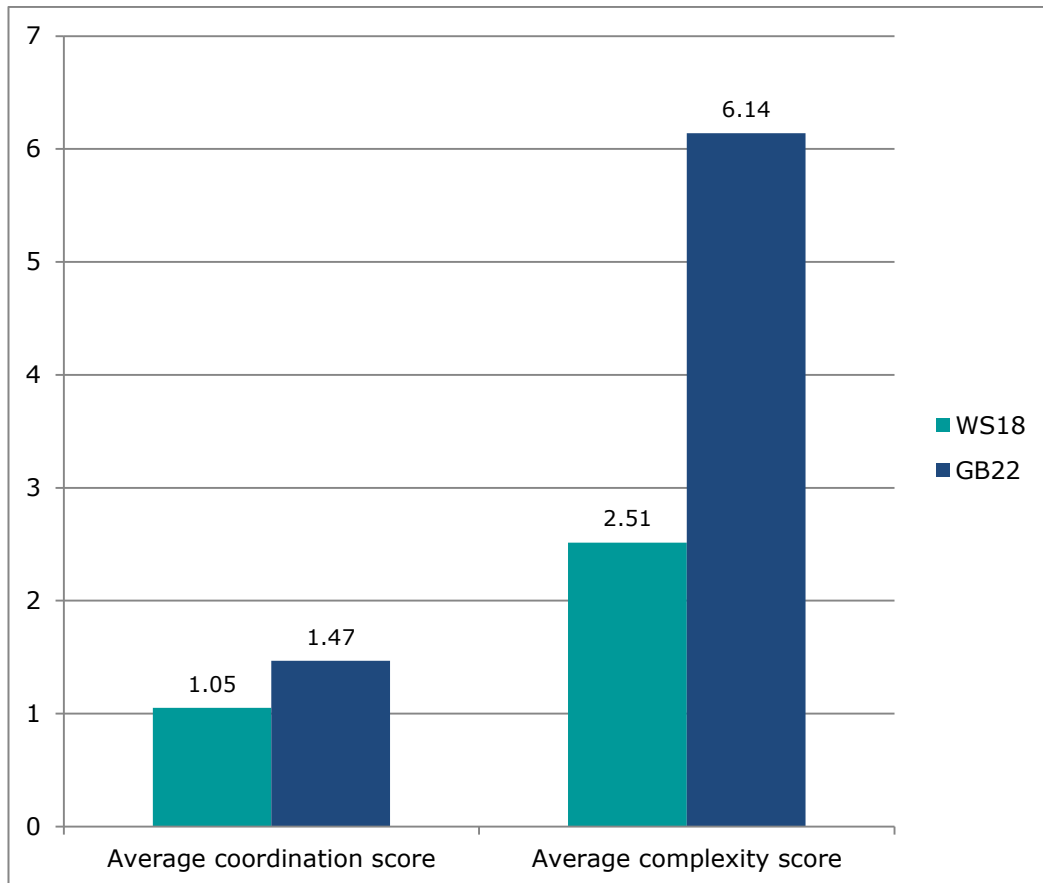
2 x dependent clauses

1 passive

Figure 6-9 shows the average coordination and complexity scores for the two chapters, derived by the methods described. While the amount of coordination in the two chapters is fairly similar, there is an extremely large contrast in the average complexity scores. This multi-factorial measure of complexity is a particularly strong discriminator between these single chapter examples of the late and early styles.

<sup>53</sup> There are commas after *sleek* and *civil* in Methuen & Co.'s 1905 edition but these do not appear in the e-text which claims to be taken from the New York Edition and are therefore not included in the *HJPC*.

**Figure 6-9 Comparison of coordination and complexity scores**



## 6.5 Washington Square Chapter 18

In section 6.5.1 I will show how data collected on the number of words in each sentence of WS18, and the coordination and complexity scores of each sentence, can be shown graphically. Such a presentation reveals both an overall pattern across the chapter for each parameter, and a number of sentences which have unusually high or low scores compared to the majority of sentences in WS18. I will argue that these overall patterns, which I have called *profiles*, correlate broadly with the narrative of the chapter as well as discourse type. In section 6.5.2 I will focus on the individual exceptional sentences (*salient* sentences) identified in the profiles, with reference to character, discourse and plot.

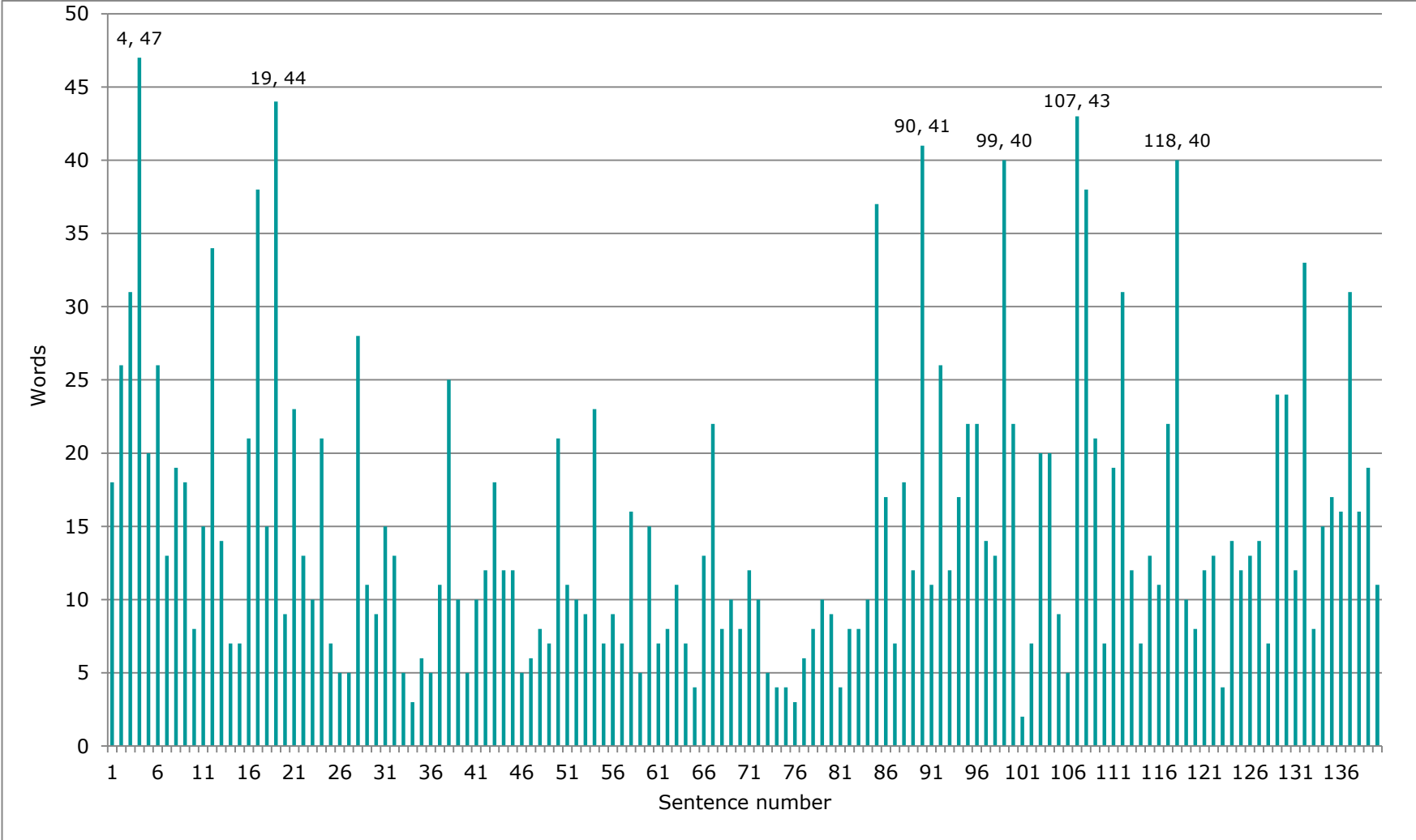
### 6.5.1 Profiles

#### 6.5.1.1 Word Profile

Figure 6-10 shows the *word profile* of WS18, that is the number of words in each sentence over the whole of WS18 in sequence, with the sentence numbers displayed along the horizontal axis. The six sentences containing 40 words or more

are marked with their statistics above their bar, showing the sentence number and number of words in that order. For instance, the longest sentence in the chapter is sentence 4 with 47 words. These six sentences display internal deviance in that they are substantially longer than the norm within this chapter of the novel. I identify them as the salient sentences of WS18 by sentence length. The cut-off point of 40 words is chosen to identify sentences which are unusual in the context of the chapter as a whole and therefore invite examination. However, this threshold is pragmatic rather than principled; the cut-off could be set lower to include more comparatively lengthy sentences. While the length chosen needs to include only a small minority of sentences which can be considered exceptional, a threshold of 35 or even 30 words might still achieve that aim. The choice of a 40-word threshold is taken here in order to provide a convenient number of the most exceptional sentences for discussion.

Figure 6-10 WS18 Number of words per sentence





The shape of WS18's word profile suggests a division into four segments. Examining the profile, a change is seen from the cluster of longer sentences at the start of the chapter after WS18:19. A series of short sentences ends at WS18:85 and the final part of the chapter, from WS18:119 onwards also features short sentences. This is illustrated in Table 6-2, which shows the percentage of sentences of 20 words or more in each segment, and the percentage of those sentences which contain direct speech. Segment 1 has many long sentences while segment 2 is characterised by mainly shorter sentences. Within segment 2 is a very marked passage, from sentence WS18:68 to WS18:84, where no sentence has more than 12 words. This marked piece of dialogue will be discussed in section 6.5.1.4. Segment 3 again has sentences which are mainly long, while in segment 4 the sentences are neither particularly long nor short.

**Table 6-2 WS18 distribution of long sentences**

Segment	Sentence range	% of sentences with 20 words or more	% of sentences which include direct speech
<b>1</b>	1-19	47.37%	10.53%
<b>2</b>	21-84	10.77%	76.92%
<b>3</b>	85-118	44.12%	61.76%
<b>4</b>	119-140	18.18%	40.91%

These four segments, which are seen in WS18's word profile, mirror stages in the plot of the chapter. Between WS18:1 and WS18:19 Catherine is sitting and considering her situation, gathering her courage to talk to her father and finally approaching him. Segment 2, from WS18:20 to WS18:84 inclusive, is the early part of the conversation between Catherine and Dr Sloper. At this stage, Catherine and her father are hopeful that there may be some compromise between them. Both hope to persuade the other of their point of view. The conversation continues with a rather different tone in segment 3 from WS18:85 onwards, where there are a much higher percentage of long sentences. WS18:85 is the point at which Catherine's obedient promise to wait "for a long time" (WS18:85) for her father's consent provokes him to the following remark which horrifies Catherine:

"Of course you can wait till I die, if you like." (WS18:85 – part)

From this point on, the discussion centres on Catherine's possible actions with reference to her father's death. Each protagonist's position becomes more

entrenched, with Catherine making the declaration which will shape the rest of her life:

“If I don’t marry before your death, I will not after,” she said. (WS18:98)

while Doctor Sloper says of Mr Townsend:

“I shall never let him speak to me again. (WS18:105)

In Segment 4 the conversation takes a slightly less combative but more emotional tone. Catherine seeks to be gentle, to find a middle way between her father’s wish for her to break away from Townsend and her own need to see him at least once more. The encounter ends, and we are left with Dr Sloper considering Catherine’s reaction with interest and amusement.

Thus the variation in sentence length can be matched to the narrative structure of the chapter. Some of this variation can be accounted for by the amount of direct speech in that segment. As shown in section 5.2.3, James distinguishes direct speech from non-speech by generally using shorter sentences for the former. However, Table 6-2 shows that this is not the whole explanation for the variation between the four segments of WS18. Segment 3 has approximately 10% less speech sentences than segment 2, yet it has approximately four times the number of sentences with 20 words or more. Segments 1 and 3 have a similar percentage of long sentences while their amount of direct speech is quite different. I suggest that segments 2 and 3 can be regarded as marked; segment 3 includes relatively long sentences to convey this particularly painful part of the encounter between Catherine and her father, which is crucial to the development of the plot of the novel as a whole. Segment 2 illustrates James’s use of short sentences for high emotion, a characteristic of his writing which will be explored in section 6.5.1.4 and section 6.6.1.4.

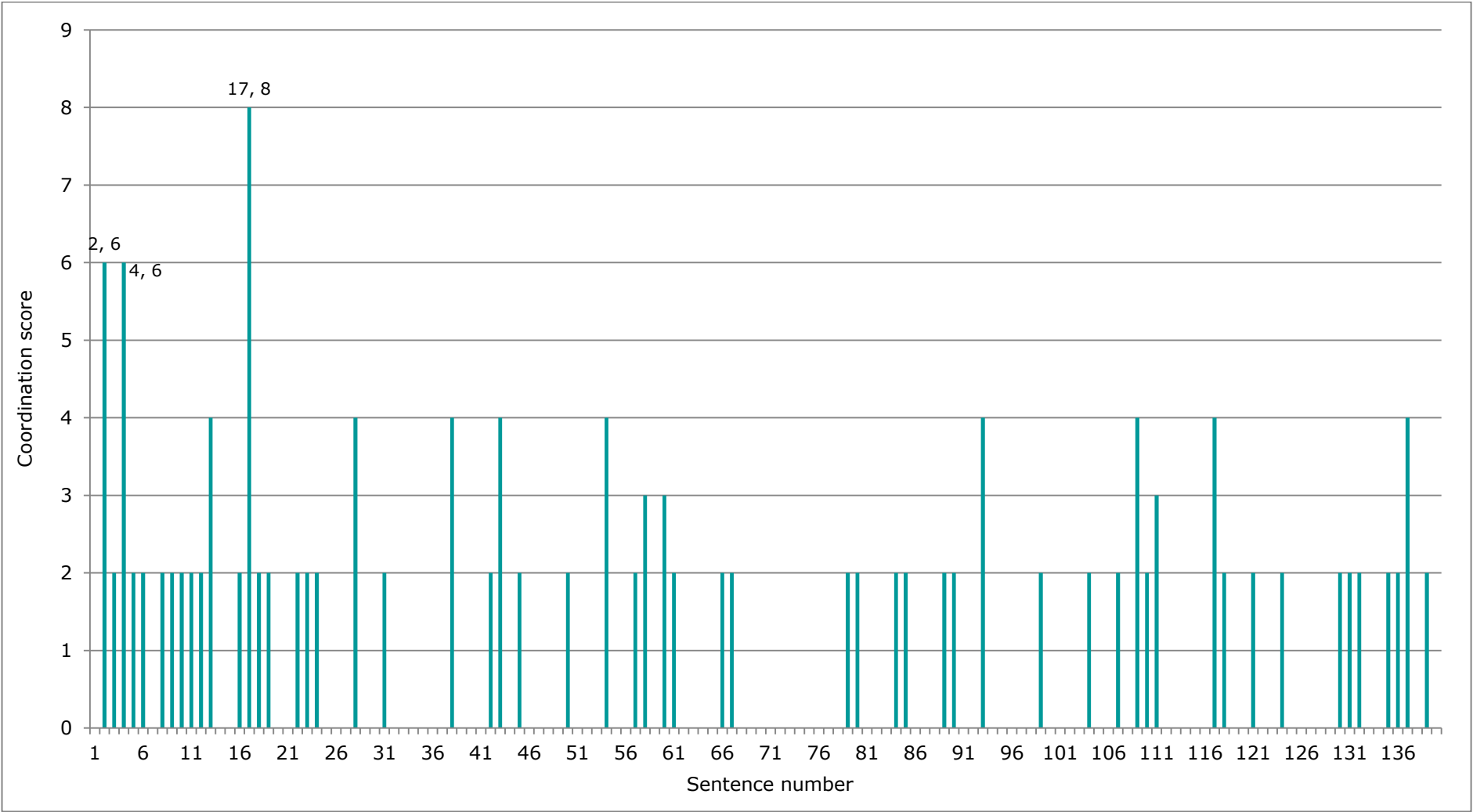
#### 6.5.1.2 Coordination profile

Figure 6-11 displays graphically the coordination score for each sentence in WS18. The *coordination profile* of WS18 is noticeably different from its word profile. Coordination is fairly evenly spread across the chapter and many sentences have no coordination at all, giving Figure 6-11 a much more even appearance than Figure 6-10. There are, however, some similarities with WS18’s word profile; there is a segment of high coordination in the early part of the chapter, although this pattern finishes here at WS18:17 rather than WS18:19. This match is unsurprising, as high coordination may contribute to long sentences. Beyond this point the coordination profile is quite homogeneous. However, the passage from WS18:68 to WS18:78, which forms part of the particularly marked part of the dialogue

identified by sentence length , is notable for being the largest area with no coordination at all. There is no area of high coordination in the final segment of the chapter, so here the coordination profile does not map onto the word profile.

There are only three sentences which stand out as salient; these are sentences 2, 4 and 17, which have coordination scores of 6 or more. They are marked on the graph with their sentence numbers and coordination scores. While there are very few sentences which are salient on account of their coordination score, combining this data with that taken from sentence and complexity scores provides a set of salient sentences for the chapter as a whole. These will be discussed in section 6.5.2.

Figure 6-11 WS18 Coordination Score per Sentence



### 6.5.1.3 Complexity profile

Figure 6-12 shows WS18's *complexity profile*. This profile is as idiosyncratic as the word and coordination profiles. Far from the relatively smooth outline of the coordination profile, the complexity profile is distinctly craggy. The four segments defined in Table 6-2 can be identified, although the boundaries of segments 1 and 2 are not exactly the same. There is an early section of more complex sentences, a central part of the chapter with mainly simple sentences, followed by a passage of greater complexity before the final simpler close to the chapter. Table 6-3 shows the data for the four segments redefined to reflect the complexity profile.

The resemblance to the word profile is unsurprising, as complexity in part reflects a greater number of clauses in the sentence. However, only four of the salient sentences in Figure 6-12 have been highlighted in Figure 6-10 and Figure 6-11: WS18:2, WS18:17, WS18:90 and WS18:118. WS18:2 and WS18:17 are highly coordinated and complex without being especially long, while WS18:90 and WS18:118 are long and highly complex, but not highly coordinated. This shows that, while there is some overlap between the measurements of sentence length and complexity, they are far from equivalent.

**Table 6-3 WS18 distribution of complex sentences**

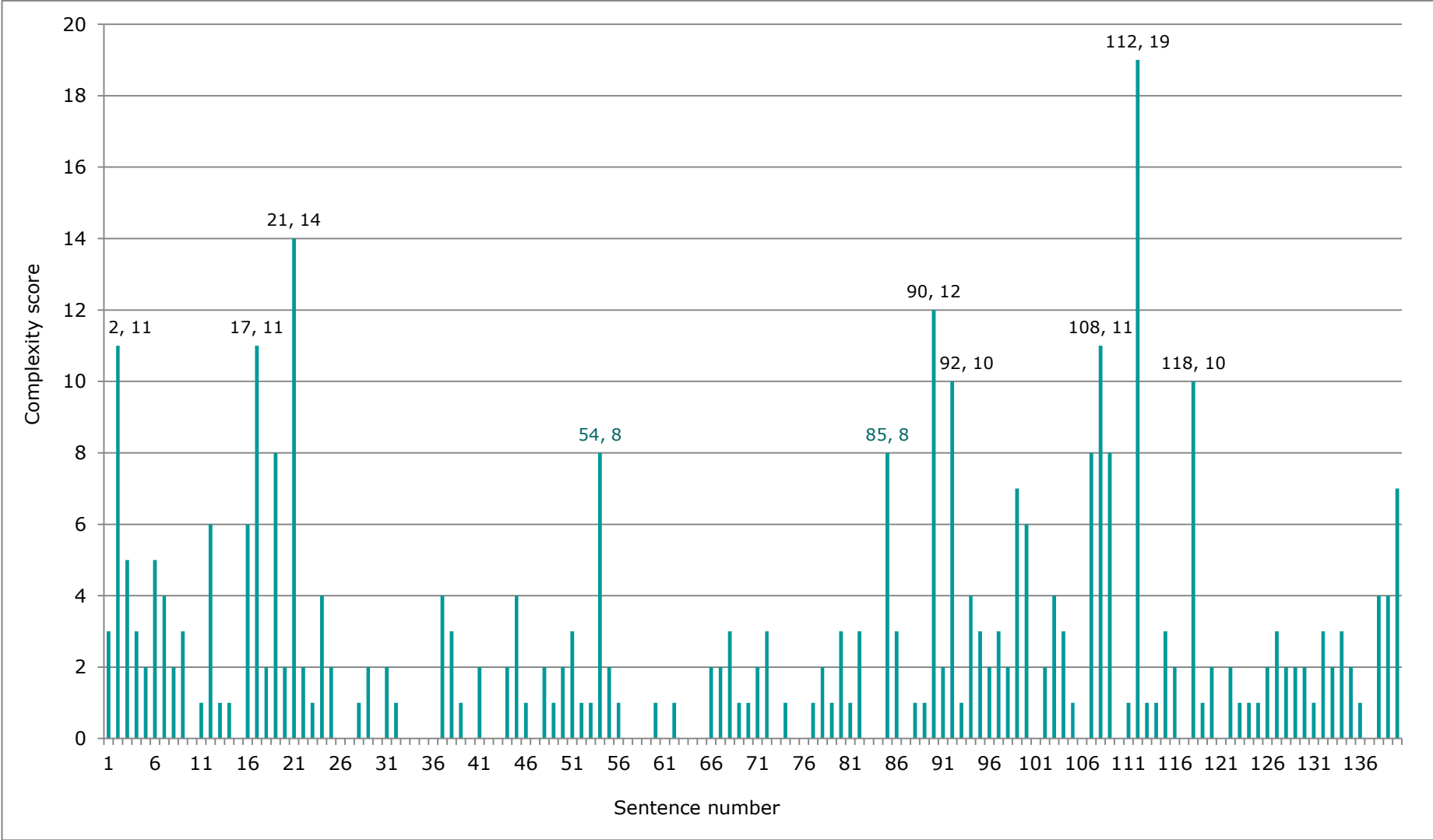
Segment	Sentence range	% of sentences with complexity score of 5 or more	% of sentences which include direct speech
<b>1</b>	1-21	38.10%	14.29%
<b>2</b>	22-89	3.17%	82.54%
<b>3</b>	90-118	31.03%	62.07%
<b>4</b>	119-140	4.55%	40.91%

Both Table 6-2 and Table 6-3 show clearly the contrast between James's use of long and complex sentences for non-speech in segment 1, and short, simple sentences for speech in segments 2, 3 and 4. As discussed in section 6.5.1.1, segment 3, however, is anomalous. More than half of its sentences contain speech and yet 44% of its sentences are relatively long and over 30% of its sentences are relatively complex. This is a key passage both in the chapter and the novel as a whole. In it, the positions of both Catherine and her father become fixed. Catherine states that she will only marry in her father's lifetime, and, by implication, with his consent. Dr Sloper says that he will not meet Mr Townsend and makes it completely

clear that he considers Catherine's suitor to be motivated by greed. The working out of this disastrous encounter is formulated in a very unusual combination of speech and complexity.

Salient sentences according to complexity are defined as those with scores of 10 or over, as shown in Figure 6-12. 8 sentences fall into that category and are marked on the graph with their sentence number and score in that order. Again, this threshold is quite arbitrary and is chosen to provide a convenient number of sentences for discussion. Two additional sentences, WS18:54 and WS18:85 are marked in turquoise; these sentences are not exceptionally complex but are notable for their relative complexity in a passage of very low complexity. They will be discussed in 6.5.1.4.

Figure 6-12 WS18 Complexity score by sentence



The three different profiles, based on sentence length, coordination and complexity scores, reflect both the plot of WS18 and the contrast between speech and non-speech, whether narration or free indirect thought. The compilation of profiles of different types has proved a useful tool for the analysis of the plot of WS18, and is particularly effective in highlighting the parts of the chapter which are most stylistically unusual and dramatically significant.

#### 6.5.1.4 Dialogue

The simple style of segment 2 of WS18 was identified by the word and complexity profiles for this chapter, though the sentences included were similar rather than identical. Here 10.77% of the sentences were longer than 20 words and 3.17% had a complexity score of more than 5. Within this segment is a passage of even more marked style. From WS18:68 to WS18:84 neither the coordination nor the complexity score rise above 3 and only one sentence is more than 10 words long (WS18:71 with 12 words). This short passage consists of Catherine and her father's conversation; the brief sentences are sometimes no more than phrases. There are far fewer interspersed sentences of narration or comment clauses in this segment, which adds to the dramatic immediacy of their exchanges. It is a painful exchange, with Doctor Sloper cold and 'grimly consistent' (WS18:130), Catherine desperately trying to find an acceptable compromise.

"What has he done - what do you know?"

"He has never done anything - he is a selfish idler."

"Oh, father, don't abuse him!" she exclaimed, pleadingly.

"I don't mean to abuse him; it would be a great mistake. You may do as you choose," he added, turning away.

"I may see him again?"

"Just as you choose."

"Will you forgive me?"

"By no means."

"It will only be for once."

"I don't know what you mean by once. You must either give him up or continue the acquaintance."

"I wish to explain - to tell him to wait."

"To wait for what?"

"Till you know him better - till you consent."



"Don't tell him any such nonsense as that. I know him well enough, and I shall never consent." (WS18:68-84)

James uses this very marked style to convey an important and emotional encounter between his two main protagonists.

Segment 2, as defined by the complexity profile and shown in Table 6-3, is interrupted by two atypical sentences: WS18:54 and WS18:85, which both have a complexity score of 8. While these sentences are not salient by my criteria, they stand out in their context. (They are shown in Figure 6-12 marked with their sentence number and complexity score in turquoise.) Sentence 54 is the one occasion in this chapter where Doctor Sloper tries to help Catherine in her difficult situation.

But it is better to be unhappy for three months and get over it than for many years and never get over it. (WS18:54)

The slightly complex structure of this sentence with its three dependent clauses, delay and extraposed subject marks it out from the sentences around it, as does the moment of gentleness and sympathy which the passage from sentence 39 onwards describes. Once Catherine fails to immediately comply with his wishes, the Doctor's patience seems to be at an end. However, here at least he suggests why the course of action he is suggesting will be in Catherine's best interests rather than merely fitting in with what Catherine sees as his prejudice against Townsend. Sentence 85 marks the ultimate failing of the attempts of father and daughter to understand each other's point of view,

"But we can wait a long time," said poor Catherine, in a tone which was meant to express the humblest conciliation, but which had upon her father's nerves the effect of an iteration not characterised by tact. (WS18:85)

and which provokes his pivotal response:

The Doctor answered, however, quietly enough: "Of course you can wait till I die, if you like." (WS18:86)

Besides this stylistically marked section of dialogue, it is possible to analyse the speech of each of the two protagonists of WS18 to see if they are differentiated by my parameters.

**Table 6-4 WS18 Speaker comparisons**

Per sentence	Catherine	Doctor Sloper
<b>Average sentence length (words)</b>	10.47	12
<b>Average coordination score</b>	0.37	0.48

<b>Average complexity score</b>	2.2188	2.22
---------------------------------	--------	------

Table 6-4 compares the sentences of the two speakers. Their average complexity scores are remarkably similar: 2.2188 for Catherine and 2.22 for Doctor Sloper. While sentence complexity overall clearly does not differentiate speakers, Dr Sloper speaks four sentences which are salient by complexity while Catherine has only two. The coordination scores show more variation. The average coordination score for the sentences in which Catherine speaks is 0.375 whereas for the Doctor it is 0.48. This difference is not very great but may serve to contrast the two characters, with Doctor Sloper more inclined to explain himself at length while Catherine uses simple, minimal constructions. The difference in the amount of coordination accounts partially, but not fully, for the difference in the average length of the two participants' speech. Catherine's sentences contain an average of 10.47 words while the Doctor's contain 12. Hence, James differentiates in a small way between Catherine and her father in the style of their speech, although it is unclear whether a reader would discern these differences either consciously or unconsciously.

### 6.5.2 Salient sentences

The sentences which have been identified as salient, or in other words foregrounded by their deviance according to the criteria explained above, are here examined in more detail. All the salient sentences are brought together and looked at in context with regard to character and plot.<sup>54</sup>

Table 6-5 shows the salient sentences identified in each of the three WS18 profiles above. All but three of the sentences (shown on white backgrounds) group together into *clusters*, as indicated by letter and the colours of the table – that is, they can be seen to form groups by being adjacent or close together within the chapter. The features which make that sentence salient are indicated in bold type: for example, sentence 2 is not unusually long with 25 words, but has high coordination and complexity scores.

---

<sup>54</sup> WS18:57 and WS18:58 are divided in the Macmillan edition and *WS*, but form one sentence in *Cornhill Magazine*. However, even as one sentence, they would not be salient by any of my criteria.

**Table 6-5 WS18 Salient sentences**

Cluster	Sentence number	No. of words	Coordination score	Complexity score
<b>A</b>	<b>2</b>	25	<b>6</b>	<b>11</b>
<b>A</b>	<b>4</b>	<b>47</b>	<b>6</b>	3
<b>B</b>	<b>17</b>	38	<b>8</b>	<b>11</b>
<b>B</b>	<b>19</b>	<b>44</b>	2	8
<b>B</b>	<b>21</b>	22	0	<b>14</b>
<b>C</b>	<b>90</b>	<b>41</b>	4	<b>12</b>
<b>C</b>	<b>92</b>	25	0	<b>10</b>
	<b>99</b>	<b>40</b>	2	7
<b>D</b>	<b>107</b>	<b>43</b>	2	8
<b>D</b>	<b>108</b>	37	0	<b>11</b>
	<b>112</b>	30	0	<b>19</b>
	<b>118</b>	<b>40</b>	2	<b>10</b>

The grouping effect points to clusters of sentences where James’s style is particularly marked by sentence length, high coordination or complexity but a similar clustering is found if sentences are chosen randomly. Therefore, while grouping the salient sentences together in clusters is useful for discussing their function, it is important not to put too much weight on the clustering effect. There is also a danger of a circular argument here. The sentences have been identified quantitatively as salient so, when evaluating their role in the chapter, there may be a predisposition to consider them important. This cannot be entirely ruled out. Intuitive critical judgement will always form a part of a stylistic analysis, even when quantitative data is used. Rigour is achieved by identifying elements to discuss objectively and by being aware of the danger of circularity.

The sentences of cluster A, WS18:2 and WS18:4, are foregrounded by their unusual complexity and length respectively, at the opening of the chapter. (WS18:3 falls outside my parameters for salient sentences but is still one of the 14 longest sentences in the chapter.)

Her aunt seemed to her aggressive and foolish, and to see it so clearly – to judge Mrs. Penniman so positively – made her feel old and grave. (WS18:2)

She had an immense respect for her father, and she felt that to displease him would be a misdemeanour analogous to an act of profanity in a great temple; but her purpose had slowly ripened, and she believed that her prayers had purified it of its violence. (WS18:4)

In this passage Catherine is sitting alone, musing on her situation. The two salient sentences are both free indirect thought, and represent Catherine's views on her aunt and father, respectively. Catherine does not usually make judgments on others, having been content to be guided by her elders, but she is beginning to feel that she has ideas and desires of her own. It is an essential point in the development of both her character and the plot that she is beginning to think for herself. The effect of these long sentences is to slow the reader's progress so that this development is more likely to be noticed. This slowing in the progress of the chapter is a stylistic counterpart to Catherine's gradual realisation of her position. We are told that this period of reflection took "more than an hour" (WS18:1) and there are repeated references to the slow passage of time as Catherine's tension builds in anticipation of the encounter with her father.

Cluster B, that is WS18:17, WS18:19 and WS18:21, at the junction of segments 1 and 2 of the chapter, portray the moment when Catherine finally plucks up the courage to go and see her father in his study, first pausing nervously in the doorway, and then finally broaching the topic of her relationship with Mr Townsend.

Her father, who was in his dressing-gown and slippers, had been busy at his writing-table, and after looking at her for some moments, and waiting for her to speak, he went and seated himself at his papers again. (WS18:17)

She remained near the door, with her heart thumping beneath her bodice; and she was very glad that his back was turned, for it seemed to her that she could more easily address herself to this portion of his person than to his face. (WS18:19)

"You told me that if I should have anything more to say about Mr. Townsend you would be glad to listen to it." (WS18:21)

The first two sentences are narrative but give Catherine's point of view; quite literally, we are given Catherine's view of her father sitting down so that she sees his back. The third sentence is Catherine's broaching of the topic of her relationship with Mr Townsend. There is again a sense of time running slowly, so much so that Dr Sloper, after waiting for Catherine to speak, returns to his desk to continue writing. The first two sentences also set the tone of the encounter to come. Dr Sloper, while not openly antagonistic, does nothing to help Catherine but rather turns his back on her. Catherine is fearful, and scarcely able to face her father. WS18:19 cannot be classified as either narration or thought. The first part is narrative but after the semicolon, Catherine's state of mind, and her thought process, is described with a comic view of her timidity. It is a moment of decision

and suspense in the chapter, and the reader pauses over the long sentence with Catherine before the vital discussion begins. This effect is heightened by the semi-colon, where the discourse type changes. WS18:21, the third sentence in this group, is anomalous in being notably complex (it is the second most complex sentence in the chapter) even though it is direct speech. There are four direct speech sentences in the chapter which are salient because of complexity but this is the only one which is spoken by Catherine. This suggests that Catherine may have been practising this opening remark beforehand. It certainly highlights the beginning of Catherine's attempt to come to terms with her father. Her hesitancy is shown in the fact that she does not actually give anything away about her position in this first remark. The complexity of the sentence is largely due to the long delay between the subordinator *that* and the clause it introduces – *you would be glad to listen to it*. This mirrors Catherine's reluctance to actually say what she wants, even though she has finally managed to get into the room and to start speaking.

Thus this group of salient sentences is important both in the narrative, and because it defines the atmosphere during the dialogue, with Dr Sloper's outward calm and Catherine's mixture of determination and nervousness.

The two sentences which most of all represent turning points in the plot are not, however, marked in any way. At WS18:86 Doctor Sloper introduces the subject of his death, which would allow Catherine to marry without considering his consent. The narrator notes that this is said 'quietly enough' (WS18:86) and the sentence is not salient by my parameters. Similarly, Catherine's definitive and ultimately binding statement -

"If I don't marry before your death, I will not after," she said. (WS18:98)

- is simply put, with nothing to mark it stylistically. The sentences are neither unusually long, co-ordinated or complex, nor are they foregrounded by their simplicity. The protagonists do not announce them dramatically, and it is likely that, within the drama, they themselves do not recognise them at the time for the sticking points they become.

With the two sentences of cluster C, which open segment 3, Dr Sloper begins to try to assert his will more forcefully, perhaps because he has 'enjoyed the point he had made' (WS18:89) by describing Catherine as looking forward to his death so that she can marry (and by implication, inherit his wealth).

It came to Catherine with the force - or rather with the vague impressiveness - of a logical axiom which it was not in her province to controvert; and yet, though it was a scientific truth, she felt wholly unable to accept it. (WS18:90)

"Give me a proof of it, then; for it is beyond a question that by engaging yourself to Morris Townsend you simply wait for my death." (WS18:92)

The length and complexity of WS18:90 mirror that of WS18:21. Though the former represents Catherine's speech and the latter her thought, both have a strong element of delay. This may, as suggested above, convey a pause while Catherine struggles to know how to proceed. It may also suggest confusion, Catherine's thoughts being muddled as the word order of the sentence is also disarranged. Dr Sloper's response at WS18:92 is complex; he is bearing down on Catherine with the sophistication of his argument, while she flounders.

WS18:99 stands alone among the salient sentences. It is a long, but not unduly complex or coordinated sentence.

To her father, it must be admitted, this seemed only another epigram; and as obstinacy, in unaccomplished minds, does not usually select such a mode of expression, he was the more surprised at this wanton play of a fixed idea. (WS18:99)

Set here in the context of the non-salient sentences which surround it, it is a strange and tragic response to Catherine's defining statement, which she thinks 'an inspiration' (WS18:97). Dr Sloper's thinking reveals his opinion of Catherine, which is that she lacks intelligence and is displaying obstinacy, if not flippancy. His thought process is convoluted and this is portrayed by the complexity of this lengthy sentence, which reveals how far apart father and daughter are in their understanding and feelings.

Cluster D marks Catherine's return to thought in response to her father's continued emphatic rejection of any possibility of compromise.

Catherine gave a long, low sigh; she tried to stifle it, for she had made up her mind that it was wrong to make a parade of her trouble, and to endeavour to act upon her father by the meretricious aid of emotion. (WS18:107)

Indeed, she even thought it wrong - in the sense of being inconsiderate - to attempt to act upon his feelings at all; her part was to effect some gentle, gradual change in his intellectual perception of poor Morris's character. (WS18:108)

Both sentences are quite long, though only WS18:107 falls into the category of salient sentences for that reason. Once again the effect is to slow the action after the preceding brief verbal exchanges, and to highlight the contrast between Catherine's gentle, submissive conciliatory approach and her father's intransigence. This contrast is further emphasised by the next salient sentence, which follows quite closely and is the most complex sentence of the chapter.

"There is one thing you can tell Mr. Townsend, when you see him again," he said: "that if you marry without my consent, I don't leave you a farthing of money. (WS18:112)

Dr Sloper considers this the last, devastating blow to Catherine's hopes. This is a misjudgement: Catherine has no intention of marrying against her father's wishes, and she is far from accepting that Mr Townsend is mercenary. The sentence is a high point of complexity and reflects the complexity of Dr Sloper's mind, while exposing his lack of empathy.

Finally, sentence WS18:118, which is both long and complex,

It won't be polite - it will express irritation; and I shall be glad of that, as it will put me in the right; unless, indeed - which is perfectly possible - you should like him the better for being rude to you." (WS18:118)

This sentence is Dr Sloper's third complex speech sentence. It reflects a new change of tone in the conversation. Dr Sloper has just laughed at his daughter's naivety about her inheritance and WS18:118 also ridicules Catherine. This sentence ends the second portion of the confrontation between Catherine and her father; the chapter ends without any more salient sentences and with Catherine behaving as submissively as she can without actually agreeing to spurn Mr Townsend.

The salient sentences, which were identified quantitatively according to their length, coordination and/or complexity, have proved to be key sentences in the unfolding of the plot of the chapter, and the novel as a whole, and in elucidating the characteristics of the two protagonists. While it would be possible to make an argument for various other sentences being important, such as the two quiet declarations of position by each protagonist seen in WS18:98 and WS18:105, this method has proved a very useful tool for navigating and examining this emotional clash around which the rest of the plot of this novel will turn.

## **6.6 *The Golden Bowl Chapter 22***

Chapter 22 of *The Golden Bowl* (GB22) is much longer than Chapter 18 of *Washington Square*, with more coordination and complexity. Nevertheless it can be analysed in the same way, identifying both profiles and salient sentences. Both of these analyses reveal aspects of plot and character, as well as the contrast James uses to differentiate speech and narrative.

The use of free indirect thought is particularly central in GB22 because James himself describes each of the two books of *The Golden Bowl*, *The Prince* and *The Princess*, as being conveyed mainly through the consciousness of their eponymous heroes.

The Prince, in the first half of the book, virtually sees and knows and makes out, virtually represents to himself everything that concerns us – very nearly (though he doesn't speak in the first person) after the fashion of other reporters and critics of other situations. (James, 2000 [Original work published 1904], p. XVIII)

James explains that this reporter-like function does not diminish the Prince's complete role as an actor in the drama, and also it is his sensitivity, with his consciousness which is 'highly susceptible of registration', which makes him such a good medium through which the first book can unfold. Chapter 22 is two chapters from the end of *The Prince*.

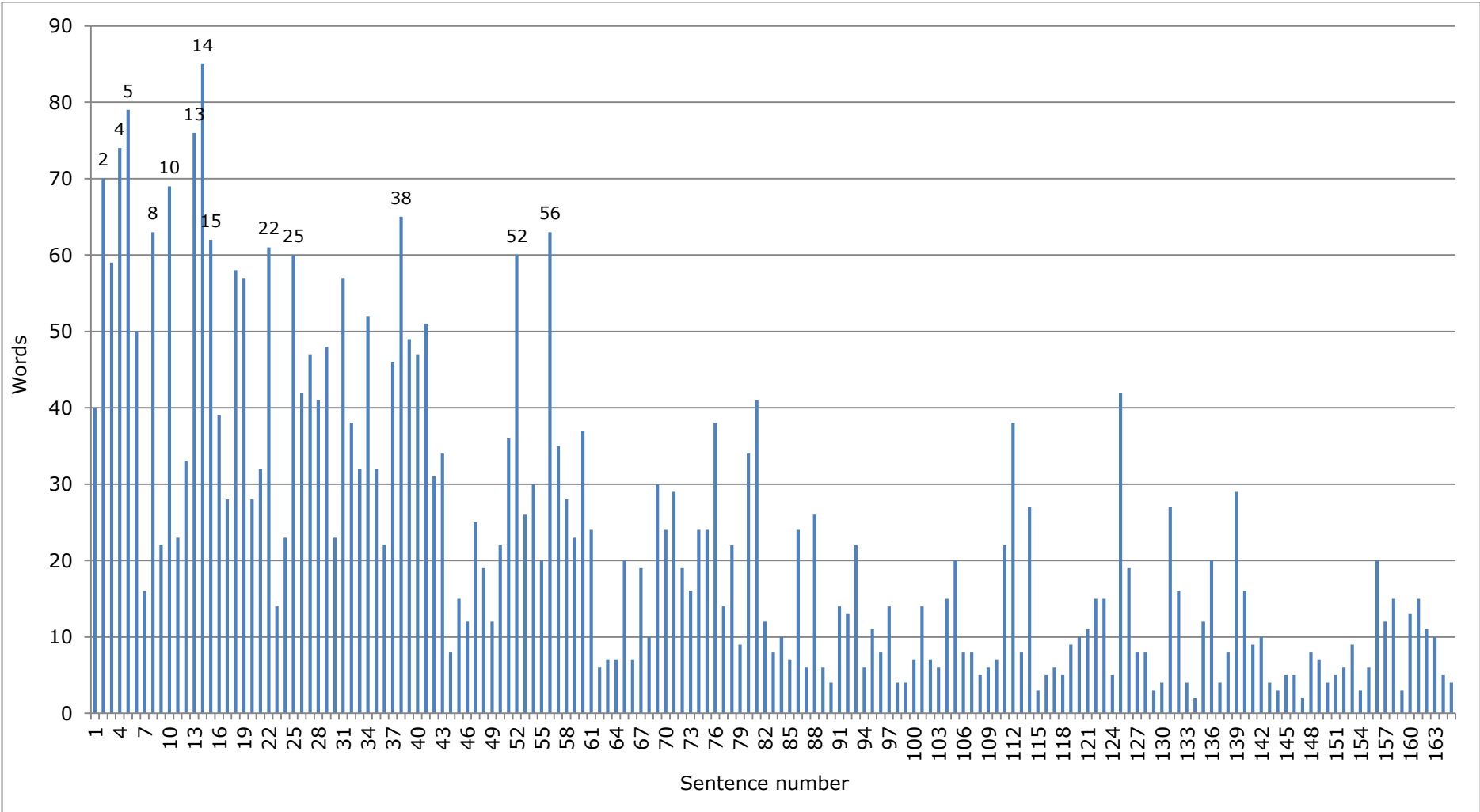
## 6.6.1 Profiles

### 6.6.1.1 Word profile

Figure 6-13 shows the word profile of GB22, that is the number of words per sentence across the 165 sentences of the chapter. The greater number of sentences in the chapter makes it more crowded than the equivalent WS18 graph, Figure 6-10. For this reason, salient sentences are marked with their sentence number and Table 6-6 is provided showing these sentences with their number of words. The salient sentences, which display internal deviance by their unusual length, are 60 words or longer. As with Figure 6-10, this threshold is chosen for pragmatic reasons, providing a convenient number of sentences for discussion.



Figure 6-13 GB22 Number of words per sentence



**Table 6-6 GB22 Longest sentences**

Sentence number	Number of words
<b>2</b>	70
<b>4</b>	74
<b>5</b>	79
<b>8</b>	63
<b>10</b>	69
<b>13</b>	76
<b>14</b>	85
<b>15</b>	62
<b>22</b>	61
<b>25</b>	60
<b>38</b>	65
<b>52</b>	60
<b>56</b>	63

In contrast with WS18's four segments, GB22 divides into two main segments which are aligned approximately with the circumstances described in the chapter. The division of the chapter is illustrated in Table 6-7. The early part of the chapter, from GB22:1 to GB22:56 inclusive, is characterised by narration, though from the Prince's viewpoint, and the Prince's free indirect thought as he strolls along the beautiful terrace of the country house at Matcham. The sentences are extremely long and only just over 7% of them contain direct speech. In segment 2, from GB22:57 onwards, Charlotte and Amerigo are mainly in conversation. The sentences are notably shorter and over 74% contain direct speech. As seen in section 5.2.3, James tends to represent speech in relatively simple, and therefore generally shorter, sentences. Within these two segments are subsegments, which contain shorter sentences than their respective segments as a whole. Segment 1a, from GB22:42 to GB22:51 inclusive, includes the Prince's brief conversation with Charlotte as she leans down from her window. Only 30% of its sentences have 30 words or more, 40% of which contain direct speech. However, segment 2a is anomalous. None of its sentences contain more than 30 words but its percentage of direct speech is slightly less than segment 2 overall. Unless this is a chance

variation, an alternative explanation to discourse differentiation is needed to explain these figures. Segment 2a will be discussed further in section 6.6.1.4.

**Table 6-7 GB22 Distribution of long sentences**

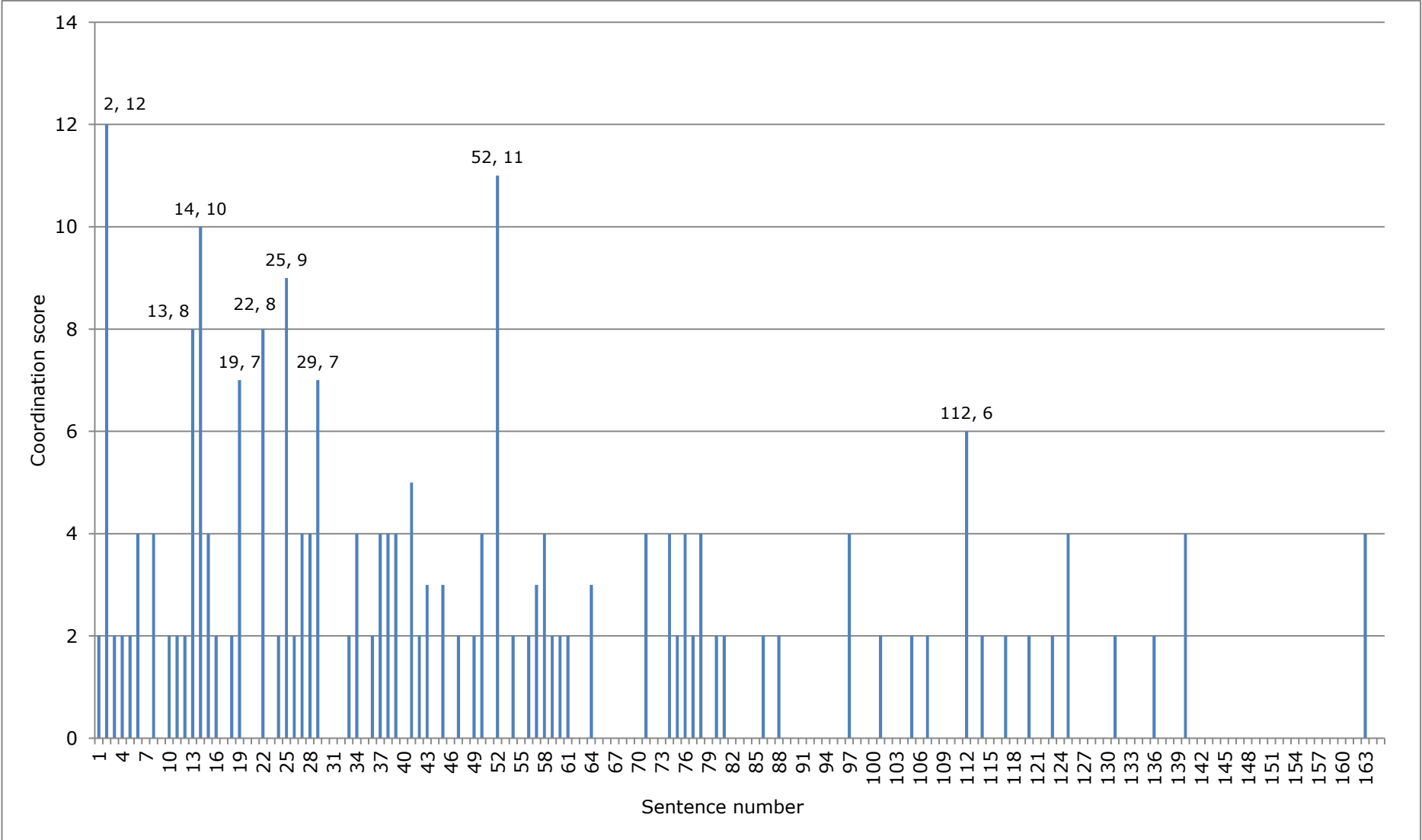
Segment	Sentence range	% of sentences with 30 words or more	% of sentences which include direct speech
<b>1</b>	1-56	67.86%	7.14%
<b>1a</b>	42-51	30%	40%
<b>2</b>	57-165	7.41%	74.31%
<b>2a</b>	82-111	0%	73.33%

### 6.6.1.2 Coordination

Figure 6-14 shows the coordination score per sentence in GB22. It includes a larger number of salient sentences than the corresponding graph in WS18 (Figure 6-11). Salient sentences are defined here as those sentences with a coordination score of 6 or above. They are marked on the graph with their sentence number followed by their coordination score. As might be expected, the most highly coordinated sentences are overwhelmingly in segment 1 of the chapter which comprises narrative and indirect thought representation; there is only one sentence which is salient by coordination in segment 2 (GB22:112). Naturally many of the sentences which have an unusual amount of coordination are also those which have already been discussed as unusually long. A great deal of coordination will often create an unusually long sentence.

However, this is not always the case. While sentences GB22:2, GB22:13, GB22:14, GB22:22 and GB22:25 are shown as salient in both Figure 6-13 and Figure 6-14, sentences GB22:19 and GB22:29 are highly coordinated but not particularly long. And GB22:52, the second most coordinated sentence in the chapter, which stands out between the first words which Charlotte and Amerigo exchange and their longer conversation, is at the lowest point of my salient sentences by sentence length at 60 words. The salient sentences will be discussed in section 6.6.2.

Figure 6-14 GB22 Coordination score per sentence



### 6.6.1.3 Complexity

Figure 6-15 shows the complexity profile of GB22. Salient sentences are those with a complexity score greater than 20 and are marked with their sentence number and score. Two sentences are marked in blue; these are unusually complex in their context, though their scores are less than 20. These two sentences are discussed in section 6.6.1.4. The contrast between the high peaks of the early representation of thought and the lower ones of the long conversation between Amerigo and Charlotte is the clearest in this profile. (Figure 6-15 shows the salient sentences with their sentence number followed by their complexity score displayed on each bar.)

Table 6-8 shows the division of GB22 into two main segments which are similar to those in Table 6-7, and two sub-segments, the second of which is quite different from sub-section 1a in Table 6-7. Segments 1 and 2 show very contrasting levels of complexity, with no sentence with a complexity score over 12 in Segment 2. Segment 1b, a passage of simpler sentences within the highly complex segment 1, centres around Amerigo's conversation with Charlotte when she is on her balcony, as does segment 1a. However, segment 1b includes a slightly wider span of sentences than segment 1a; it seems that the Prince's train of thought becomes less complex as it turns to Charlotte, describing her before she speaks, and reflecting on the scene in the context of their relationship. The conversation from the balcony itself runs only from GB22:44 to GB22:50. Segment 2b, while having a marginally lower percentage of sentences containing direct speech, is marked by extremely simple sentences; the complexity scores are zero throughout. This segment will be discussed in section 6.6.1.4.

However, there is no absolute link between complexity and discourse. Even in segment 1 outside 1b, some sentences are very simple. Rather segment 1 contains many sentences of high complexity interspersed with a few simple sentences. Without this, the novel would become almost unreadable; the simpler sentences temper the most baroque flights of James's late style.

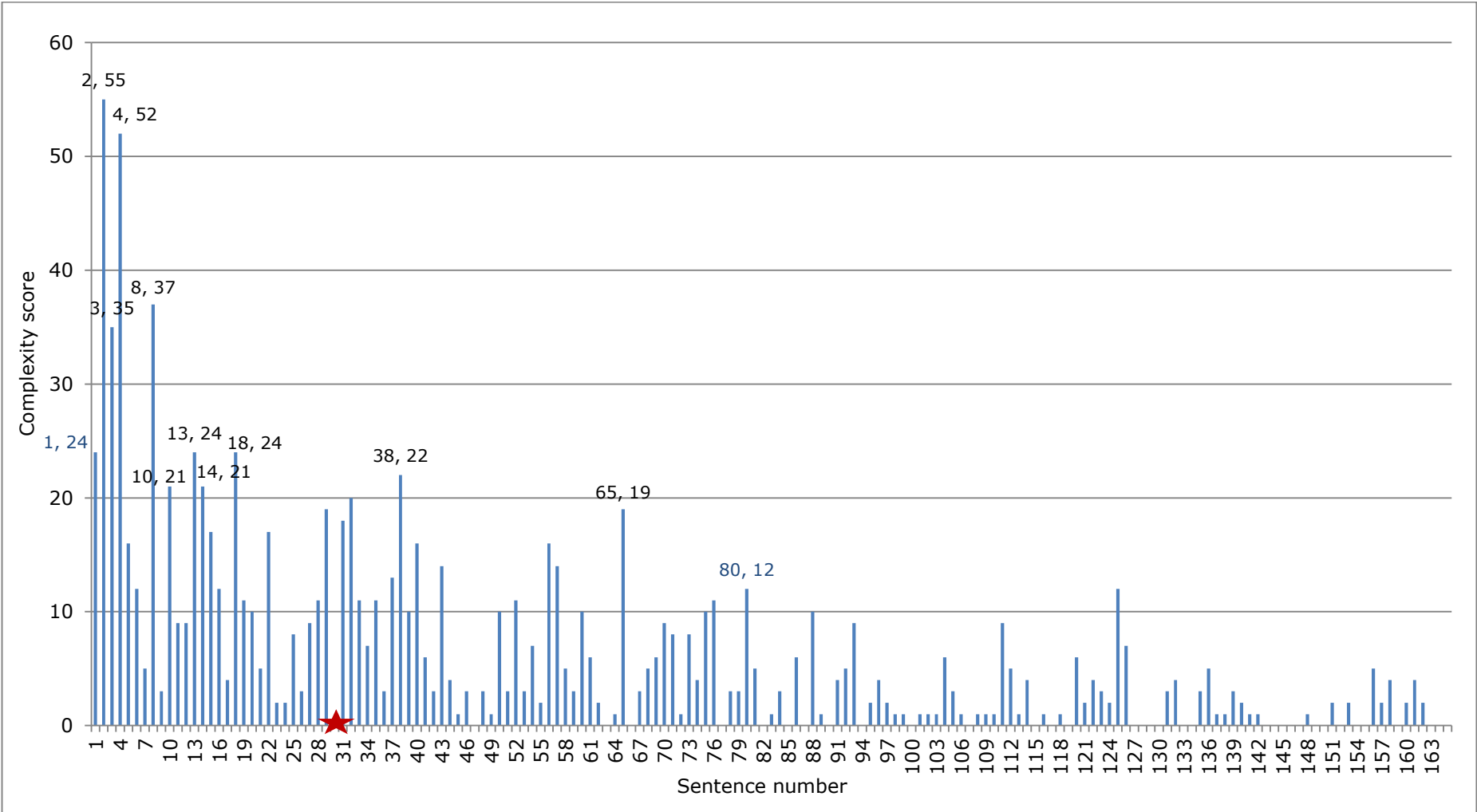
**Table 6-8 GB22 Distribution of complex sentences**

Segment	Sentence range	% of sentences with complexity score of 10 or more	% of sentences which include direct speech
<b>1</b>	1-65	50.77%	6.15%
<b>1b</b>	41-55	20.00%	26.67%

<b>2</b>	66-165	5.00%	79.00%
<b>2b</b>	127-165	0%	76.92%

Sentences can also be marked as deviant by their unusual simplicity. GB22:30, which is marked on Figure 6-15 by a red star, has a complexity score of zero even though it is free indirect thought and lies between two sentences with complexity scores of 19 and 18 respectively. GB22:30 has been counted as salient for that reason, and will be discussed in section 6.6.2.

Figure 6-15 GB22 Complexity score per sentence



#### 6.6.1.4 Dialogue

I will now turn to a closer analysis of some of the most stylistically marked parts of the dialogue within GB22. Figure 6-13, Figure 6-14 and Figure 6-15 show clearly the contrast between segment 1 of GB22, which consists largely of narrative and free indirect thought, and segment 2, which contains the dialogue between Charlotte and Amerigo. However, while the dialogue in GB22 is mainly in segment 2, i.e. from GB22:61 onwards, there is a short conversation between the Prince and Charlotte in segment 1a, from GB22:39 to GB22:55. Segment 1a contains some sentences which do not include speech, but they are generally short, with low coordination and complexity, resembling the speech sentences more than the long, complex and highly coordinated sentences of the Prince's thought process in segment 1 as a whole.

There are some sentences in GB22 which, while not salient by my parameters, stand out by their length, coordination or complexity with reference to their context within the dialogue. These, then, are also likely to draw the attention of the reader as deviant in their context, and stand out as foregrounded when examined in the context of plot and characterisation. For example, GB22:80 and 81 are both quite long, with 34 and 41 words, respectively. Their coordination scores are low at 2, but the complexity score of GB22:80 is unusually high at 12, while GB22:81 is a more average 5. (GB22:80 is one of only two sentences which contain direct speech and have a complexity score as high as 12. The other is GB22:125, discussed below.) GB22:80 is also one of only two sentences more than 40 words long in the second segment of the chapter. All these characteristics and their context mark these sentences out for analysis. GB22:80 and 81 immediately precede the markedly short group of sentences identified as segment 2a in Table 6-7. Amerigo and Charlotte are discussing the golden bowl, which they saw in the early part of the novel and which has not been explicitly mentioned since. They found the bowl when they were shopping together for a wedding gift for Charlotte to give the Prince and Maggie Verver. Amerigo had rejected it as a possible gift because it had a crack in it. Austen-Smith claims that the discussion of the bowl, to which the reader may think the Prince is referring, but of which he apparently has to be reminded by Charlotte, shows that their bond is not as strong as they think.

At the moment in the novel where the Prince and Charlotte supposedly enjoy perfect understanding, they really do not understand each other at all. The bowl operates here as a failed Eucharist, a counterfeit symbol of the perfect union it seems at first to capture. Charlotte and Amerigo may exchange metaphors that accidentally resemble one another, but they have no common referent. (Austen-Smith, 2004, p. 58)



Once Amerigo is reminded of the bowl and the Jewish trader who was selling it, his remark in GB22:80 (shown below) may have an element of irony in referring to the cracked bowl, which he has already rejected, as being kept for Charlotte.

“Well, you also, no doubt, made a great impression on him, and I dare say that if you were to go back to him you’d find he has been keeping that treasure for you. (GB22:80)

There is an implication here that the ‘treasure’ which is available to Charlotte in her relationship with the Prince can only be this cracked one, which he would not accept as a gift to his wife. The subsequent sentence, which is prominent by its length, seems to go further, warning Charlotte while they are actually planning their day together.

But as to cracks,” the Prince went on - “what did you tell me the other day you prettily call them in English? ‘rifts within the lute’? - risk them as much as you like for yourself, but don't risk them for me.” (GB22:81)

‘Rifts within the lute’ is a quotation from Tennyson’s ‘Merlin and Vivien’ which forms part of his long Arthurian poem *Idylls of the King* published in its final form in 1885.<sup>55</sup> (Tennyson, 1983 [Original work published in various versions from 1842-1899], p. 152) The metaphorical use of ‘cracks’ is introduced by Amerigo, who hopes that Charlotte is not suggesting that the current ‘occasion’ is cracked. Not waiting for an answer he suggests that Charlotte may “think too much of ‘cracks’” and fear them. (GB22:77) The meaning is somewhat obscure but presumably he is persuading Charlotte to allow their liaison to be consummated. After discussing the Jew who was selling the bowl, the Prince seems to backtrack when he tells Charlotte in GB22:81 to risk cracks ‘as much as you like for yourself, but don’t risk them for me’. If Charlotte decides to go ahead, she must take responsibility for the decision herself. For himself, the Prince feels the omens are good.

The implications of the quoted phrase ‘rifts within the lute’ give another dimension to the conversation. The quotation is spoken by a character called Vivien in Tennyson’s poem but she is quoting a song sung by Lancelot. The song is about ‘Faith and unfaith’ in love, declaring that total faith is required. It is ‘unfaith’ which is ‘the little rift within the lute’ and its effect is disastrous; ‘rotting inward [it] slowly moulders all’ (Tennyson, 1983 [Original work published in various versions from 1842-1899], p. 152). There are layers of irony here. If the Prince is implying a comparison between Charlotte and Vivien, it is not a flattering one. Vivien is one of the villains of the Tennyson’s Arthurian story, coming to the court specifically for her own ends, sowing disaster when the opportunity arises with Balin and Balan,

---

<sup>55</sup> I am indebted to Philip Horne for the identification of this reference. His paper ‘Henry James among the Poets’ details James’s admiration of Tennyson - “the poet I had earliest known and best loved” (Horne, 2005, p. 72).

and settling on the seduction of Merlin simply as the most likely route to gain power. It is Vivien who finally seals Merlin into a tree. James indicates to us here a possible view of Charlotte, at least in the eyes of her lover, as a self-serving seductress and adulteress. The Prince may be making this reference consciously or unconsciously but either way it is a startlingly discordant note in the romantic scene on the terrace. However, the use of a quotation which is really Lancelot's is also important. Lancelot is an adulterer himself, betraying his friend and King with Queen Guinevere. It is difficult to imagine that this is a conscious comparison from Amerigo. Rather the unconscious equivalence between himself and the infamous adulterer brings out the ambivalent quality of this relationship which involves so much betrayal. GB22:80 and GB22:81 drew attention by their unusual length and complexity as part of the dialogue of GB22, and particularly because they are immediately followed by a group of short sentences. On examination, these sentences convey a warning in veiled words to Charlotte, and to the reader, that the Prince is by no means committed to a devoted love affair, even though he immediately moves the liaison forward.

As in WS18, emotionally charged dialogue in GB22 is conveyed in very short, and particularly by very simple, sentences. This is seen in the stylistically marked dialogue contained in segments 2a and 2b, identified in Table 6-7 and Table 6-8. Segment 2a, from GB22:82 to GB22:111, was identified from the word profile because these sentences all contain less than 30 words. This is unusual even in the context of segment 2's dialogue. The conversation turns here from an oblique discussion of the Prince and Charlotte's situation through the metaphor of the cracked golden bowl to something more personal and direct. Now the interchange becomes more intense and each lover takes a step into a further commitment to their liaison, although still somewhat obliquely.

"I go, as you know, by my superstitions. And that's why," he said, "I know where we are. They're every one today on our side."

Resting on the parapet toward the great view she was silent a little, and he saw the next moment that her eyes were closed. "I go but by one thing." Her hand was on the sun-warmed stone; so that, turned as they were away from the house, he put his own upon it and covered it. "I go by YOU," she said. "I go by you." (GB22:83-90)

Segment 2b, from GB22:127 to GB22:165, is a similarly intense passage but is identified mainly from the complexity profile. Here no sentence has a complexity score of more than zero, only four sentences contain any coordination at all and no coordination score is higher than four, and only two sentences are longer than twenty words. This unusually simple passage is where Charlotte makes it clear that she is not only ready to go with Amerigo but has actually made all the arrangements necessary for them both to get away alone. In this intense exchange

the conspiracy is confirmed and their intimacy is brought into the open, both to each other and to the reader. While the words are not explicit, there is an atmosphere of sexual tension, of the two lovers at last taking real steps to be alone together. In both segments 2a and 2b, James's method of using short, simple exchanges to express intensity has clearly survived from his early writing in Washington Square and is used to powerful effect here.

Two speech sentences from Charlotte stand out in this conversation. One is a salient sentence by my parameters. GB22:112 is quite long at 36 words, and is salient by virtue of its coordination score of 6. It is the most highly coordinated sentence in segment 2, though not unusually complex (complexity score 5). This is the sentence in which the whole plan to be in Gloucester alone, as arranged by Charlotte, comes together.

However, I'm sure 'Glo'ster Glo'ster' will be charming," she still added; "we shall be able easily to lunch there, and, with our luggage and our servants off our hands, we shall have at least three or four hours. (GB22:112)

Charlotte's most complex speech sentence is GB22:125, which presents her alibi for them to Lady Castledean but also to their spouses. They will ostensibly visit Gloucester Cathedral as tourists, an idea she presents ironically as their family's wish and expectation. She impresses the Prince with the extent of her complicity in the deception. The unusual complexity of the form of this sentence mirrors the complicated levels of truth and lies which will facilitate their liaison.

"Why, that we like cathedrals; that we inevitably stop to see them, or go round to take them in, whenever we've a chance'; that it's what our respective families quite expect of us and would be disappointed for us to fail of. (GB22:125)

The implications of this fabrication of an adulterous deception are emphasised when, in the next sentence, Charlotte is suddenly referred to by the narrator (but perhaps in the Prince's mind too) as 'Mrs. Verver' when she has been 'Charlotte' throughout the chapter to this point.

Thus the sentences which stand out within Charlotte's speech convey Charlotte's acceptance of an affair and her practical work for its consummation. The Prince's notable sentences, on the other hand convey a reservation and a warning – the crack in the golden bowl.

Table 6-9 compares the data for Amerigo and Charlotte's speech respectively in order to explore the possibility that James differentiates between them by the style of their speech.

**Table 6-9 GB22 Speaker comparisons**

<b>Per sentence</b>	<b>Amerigo</b>	<b>Charlotte</b>
<b>Average number of words</b>	12.28	12.38
<b>Average coordination score</b>	0.53	0.78
<b>Average complexity score</b>	2.67	2.56

Unlike WS18, there is no obvious discrepancy of intelligence and sophistication between the two characters which may lead us to expect a differentiation of complexity in their speech. In fact, as Norrman (1982) has pointed out (see section 3.3.2), James does not use clearly idiosyncratic styles for his characters; the average complexity scores of the two speakers are very similar, though not so identical as Doctor Sloper and Catherine. There is a differentiation of coordination, with Charlotte using rather more, correlating with slightly longer sentences. Using these markers of style, the reader cannot be expected to differentiate speakers, and keeping track of who is speaking is not, in fact, always an easy task. For example, in the passage where they discuss Charlotte's information about which train they should use to go to Gloucester, sentence 143 is, at first glance, difficult to interpret, and the reader has to rely on placement of speech marks to allocate the remark to Charlotte.

"You looked it up – without my having asked you?" (GB22:141)

"Ah my dear," she laughed, "I've seen you with Bradshaw! (GB22:142)

It takes Anglo-Saxon blood." (GB22:143)

"Blood'?" he echoed. (GB22:144)

"You've that of every race!" (GB22:145)

It kept her before him. (GB22:146)

"You're terrible." (GB22:147)

Well, he could put it as he liked. (GB22:148)

However, speech marks do not help with the decoding of GB22:145. The reader may assume that Amerigo is continuing to speak from GB22:144 to the return to his thought in GB22:148. However, it is also possible that Charlotte tells Amerigo that he has many races in his blood. Indeed, it is only the precedent in this chapter (and in this section of the book) of the Prince being the focalizer of all the thought presentation which suggests that it is Amerigo's thoughts which are reported in sentence 148, and not Charlotte's. Thus, while the style is simple here, as

measured by sentence length, coordination and complexity, it is not simple for the reader to negotiate the passage in terms of speaker turns, or, as will be discussed in section 6.6.2, in terms of the protagonists' feelings.

### 6.6.2 Salient sentences

Table 6-10 shows in sentence order all the salient sentences previously identified. The colour shading groups the sentences in clusters which occur close to each other. GB22:38 stands alone. The score(s) for which each sentence is deemed salient is shown in bold type.

**Table 6-10 Salient sentences**

Cluster	Sentence number	Number of words	Coordination Score	Complexity Score
A	1	40	2	<b>24</b>
A	2	<b>70</b>	<b>12</b>	<b>55</b>
A	3	59	2	<b>35</b>
A	4	<b>74</b>	2	<b>52</b>
A	5	<b>79</b>	2	16
A	8	<b>63</b>	4	<b>37</b>
B	10	<b>69</b>	2	<b>21</b>
B	13	<b>76</b>	<b>8</b>	<b>24</b>
B	14	<b>85</b>	<b>10</b>	<b>21</b>
C	18	58	2	<b>24</b>
C	19	57	<b>7</b>	11
C	22	<b>61</b>	<b>8</b>	17
C	25	<b>60</b>	<b>9</b>	8
C	29	48	<b>7</b>	19
	<b>38</b>	<b>65</b>	4	<b>22</b>
D	<b>52</b>	<b>60</b>	<b>11</b>	11
D	<b>56</b>	<b>63</b>	2	16
	<b>112</b>	38	<b>6</b>	5

These salient sentences, arranged into four clusters (A, B, C and D), bring the attention of the reader to important revelations in the plot of the novel. As this is very much a novel of thought and feeling, insights into the Prince's mind are as much important plot developments as the actual progress of his affair with Charlotte. The sentences which are marked by length, coordination and complexity often focus on these revelations.

More extensively than in WS18, GB22 starts with a group of long and complex sentences which set the scene and portray the feelings and thought process of the main protagonist. In this passage we learn not only the situation the Prince is in, but his interpretation of that situation. He considers the women in his life – Fanny, Maggie and Charlotte, and even his hostess – and the help he has had from them. His attitude is summarised in the second salient sentence of cluster A, using the metaphor of the view – ‘his extraordinarily unchallenged, his absolutely appointed and enhanced possession of it’ (GB22:2). Four out of the six sentences in cluster A have the highest complexity scores in the chapter, while sentence 2 also has the most coordination in the chapter. Sentence 5 is unusually long but otherwise unexceptional. Reading this opening passage is a slow process, requiring the negotiation of these difficult sentences and perhaps compelling the reader to notice the Prince’s attitude. Cluster A brings vividly into focus the Prince’s arrogance and his sense of great entitlement to the devotion of women. GB22:1 begins the chapter on a deceptive note of bucolic calm; only when we get to the end of GB22:2 and the phrase ‘his absolutely appointed and enhanced possession of it’ (i.e. the view from the terrace) does the jarring note of the Prince’s arrogance appear. The view, and the occasion, is presented as a work of art which he has collected, just as the Ververs’ have travelled round Europe choosing art for their collection, including, arguably, the Prince himself.

It was quite for the Prince after this as if the view had further cleared; so that the half-hour during which he strolled on the terrace and smoked - the day being lovely - overflowed with the plenitude of its particular quality.

Its general brightness was composed doubtless of many elements, but what shone out of it as if the whole place and time had been a great picture, from the hand of genius, presented to him as a prime ornament for his collection and all varnished and framed to hang up - what marked it especially for the highest appreciation was his extraordinarily unchallenged, his absolutely appointed and enhanced possession of it.

Poor Fanny Assingham's challenge amounted to nothing: one of the things he thought of while he leaned on the old marble balustrade – so like others that he knew in still more nobly-terraced Italy - was that she was squared, all-conveniently even to herself, and that, rumbling toward London with this contentment, she had become an image irrelevant to the scene. (GB22:1-3)

GB22:3, 4 and 5 reveal the Prince’s real feelings about the women in his life. He is far from romantic, seeing them as creatures to be manipulated for his convenience and enjoyment. At present he is happy because all his women are working for ‘his interest’ (GB22:5) but there is a suggestion that he could be less benevolent if that were not the case.

It further passed across him - as his imagination was, for reasons, during the time, unprecedentedly active - that he had after all gained more from women than he had ever lost by them; there appeared so, more and more,

on those mystic books that are kept, in connexion with such commerce, even by men of the loosest business habits, a balance in his favour that he could pretty well as a rule take for granted.

What were they doing at this very moment, wonderful creatures, but trying to outdo each other in his interest? - from Maggie herself, most wonderful in her way of all, to his hostess of the present hour, into whose head it had so inevitably come to keep Charlotte on, for particular reasons, and who had asked in this benevolent spirit why in the world, if not obliged, without plausibility, to hurry, her husband's son-in-law shouldn't wait over in her company. (GB22:4-5)

Cluster B, that is GB22:10, GB22:13 and GB22:14, is the next most complex cluster and two of the three sentences are also highly coordinated. It is more specific and present-focused than cluster A, although it also uses the view from the terrace at Matcham as a starting point. Amerigo seems torn between triumph at the opportunity to be with Charlotte and anger at being used as a foil by his hostess. The Prince's relationship to English society, with all its ambivalence and hypocrisy, is clarified and his feelings about his position made clear in the longest sentence of the chapter.

The Prince had the sense, all good-humouredly, of being happily chosen, and it wasn't spoiled for him even by another sense that followed in its train and with which during his life in England he had more than once had reflectively to deal: the state of being reminded how after all, as an outsider, a foreigner, and even as a mere representative husband and son-in-law, he was so irrelevant to the working of affairs that he could be bent on occasion to uses comparatively trivial. (GB22:14)

Here the Prince's bitterness, the injury to his pride of his perceived irrelevance to his wife and father-in-law, is made clear. The Prince justifies his adultery to himself in this way, 'all good-humouredly'; the situation suits his plans but there is no disregarding his discomfort, of his sense of being a mere decorative item in the Ververs' great art collection. Amerigo's feelings seem ambivalent themselves; while he welcomes the opportunity afforded him by Lady Castledean's intrigue, he is also offended by being used by her, and, he feels, by his wife and father-in-law.

The five sentences of cluster C, starting with GB22:18 and ending with GB22:29, form a group delineated mainly by coordination. It is possible that there is a link between the list-like quality of these highly coordinated sentences and the Prince's sense of experiencing and noting many of the characteristics of the English without being able to process them into an understanding of the nation. The Prince reflects on his relationship to the aristocratic English social scene in which he is involved. In this circle he is not only an outsider but also, he considers, 'among all these so often inferior people, practically held cheap and made light of' (GB22:18). As he considers his situation he tells himself that he 'could rise above' (GB22:19) the judgement of the English. He does not understand them entirely:

He knew them all, as was said, "well"; he had lived with them, stayed with them, dined, hunted, shot and done various other things with them; but the number of questions about them he couldn't have answered had much rather grown than shrunken, (Part of GB22:22)

Amerigo often finds himself 'confronted with a mere dead wall, a lapse of logic, a confirmed bewilderment' (GB22:29). However, the important characteristic of the English in his present situation is their love for compromise which has allowed Lady Castledean to keep Mr Blint with her by inviting Charlotte and Amerigo to stay on at Matcham as chaperones. This set of sentences, highlighted by their coordination, is important in establishing the Prince's attitude to his social circle. He is bemused by the English, feels that they do not accept him and retaliates by despising them while finding their odd behaviour convenient for his purposes. The Prince never loses his awareness of his superiority, perhaps because of the importance he gives to his royal blood. With this in his mind, the idyllic scene described at the beginning of the chapter takes on a different colour, so that 'the earth and the air, the light and the colour, the fields and the hills and the sky, the blue-green counties and the cold cathedrals' (GB22:25) are seen through the veil of English complacency. The reference to 'cold cathedrals' is particularly significant as a visit to Gloucester Cathedral is to be the cover story adopted by Charlotte and Amerigo, allowing them to spend time alone together. At the least, this adds irony to their ruse when Charlotte later suggests it. More than that, it may be a rather ill-omened reference.

Sentences may also be salient because of their unusually low score, as the reader's attention may be drawn to a sudden simplicity as well as complexity. Sentence 30, marked on Figure 6-15 with a red star, has a complexity score of 0 although it is free indirect thought. The sentences on either side of it have quite high complexity scores of 19 and 18 respectively. Clearly it is deviant and the foregrounding seems motivated when the sentence is examined.

And moreover above all nothing mattered, in the relation of the enclosing scene to his own consciousness, but its very most direct bearings. (GB22:30)

This is the moment where the Prince pulls himself out of his musing to think of practical matters – where is Charlotte and how can he pursue his purpose? There are other sentences with complexity scores of 0, but only 8 others are not direct speech. They are narrative sentences within the first or (mostly) the second and longer conversation. The relative brusqueness and focus of GB22:30 come as a shock after the long meanderings of Amerigo's walking and thinking. The shock is mirrored and highlighted by the sudden (relative) simplicity of the sentence, which in turns brings into focus the Prince's ability to be ruthless in the pursuit of his goals.



Sentence 38 also stands alone, but is marked by its complexity. It depicts the moment where Amerigo realises that Charlotte is ready for 'the larger step'. (GB22:38)

The larger step had been since the evening before intensely in his own mind, though he hadn't fully thought out even yet the slightly difficult detail of it; but he had had no chance, such as he needed, to speak the definite word to her, and the face she now showed affected him thereby as a notice that she had wonderfully guessed it for herself. (GB22:38)

This revelation of Charlotte not only understanding his desire to take the opportunity to consummate their affair but being ready to do so is an important addition to our knowledge of both characters. Until this point nothing has been made explicit (even to this guarded degree) about their relationship. Now it is clear that it is an affair and that both are equally engaged in moving matters forward.

Cluster D is a more dispersed grouping. Sentence 52 is Amerigo's appreciation of the distant view of Gloucester, which merges in his mind with his freedom, elation and opportunity to be with Charlotte.

This place, with its great church and its high accessibility, its towers that distinguishably signalled, its English history, its appealing type, its acknowledged interest, this place had sounded its name to him half the night through, and its name had become but another name, the pronounceable and convenient one, for that supreme sense of things which now throbbed within him. (GB22:52)

The city of Gloucester becomes consciously to the Prince a metaphor for his affair with Charlotte. Sentence 56 is the revelation of Amerigo's cynical conclusion that the boredom of his marriage has been worth it because it has bought him the freedom to have a liaison with Charlotte.

He knew why he had from the first of his marriage tried with such patience for such conformity; he knew why he had given up so much and bored himself so much; he knew why he had at any rate gone in, on the basis of all forms, on the basis of his having in a manner sold himself, for a *situation nette*. (GB22:56)

This 63-word sentence is a shock in its directness: this is the truth about Amerigo and Maggie's marriage. The phrase *situation nette*, which can be translated as a 'clear-cut situation', also has overtones of financial dealings, as the word 'nette' (a feminine form) is analogous to the English net, in the sense of net of tax, for example. The Prince's marriage is, in its essence, a financial transaction; he has 'sold himself'.

Sentence GB22:112, which is not part of a cluster, was discussed in the context of Charlotte's dialogue in section 6.6.1.4.

## 6.7 Conclusion

This detailed analysis of the central chapters of *Washington Square* and *The Golden Bowl* illustrates the strengths and weaknesses of a corpus-based quantitative methodology. The different profiles, derived from data on sentence length, coordination and complexity, provide a skeleton of the chapters' plots and discourse types. In WS18, Catherine's assessment of her situation, her hesitant approach to her father's study and some key points in their conversation are all revealed by foregrounded, high-scoring sentences while their intense, emotional conversation appears in short, simple exchanges. Similarly, the shape of GB22 emerges, with the long and extremely complex sentences of the Prince's musing over his relationships with women and with English society giving way to his conversation with Charlotte. Again intense emotion, though of a very different sort, is depicted with simple sentence construction. The use of salient sentences is also a very useful tool for [identifying the foregrounding which](#) highlights some of the key moments and ideas of each chapter. The possibility that Catherine has researched her opening remark to her father, for instance, is suggested by the unusual complexity of WS18:21, and the complexity of Charlotte's speech in GB22:125 complements its content - the lie she and the Prince will tell to Maggie and Mr Verver.

However, while many of the salient sentences are clearly extremely important in the story of their respective chapters, there are important sentences which are not salient by my measures. Some sentences will be foregrounded in other ways. For instance, Charlotte is referred to as Mrs Verver only twice in the chapter, once when she appears on the terrace and again as she explains their alibi for visiting Gloucester. As this part of *The Golden Bowl* is overwhelmingly from the Prince's point of view, it may be that the change of name is foregrounded to suggest that the Prince becomes more aware that Charlotte is his father-in-law's wife at those points in the narrative. Foregrounding can be achieved in many ways.

With this caveat, the methods developed for, and employed in, this chapter have proved a successful way of using quantitative corpus data to guide a rigorous stylistic reading of literary texts.

# Chapter 7 Conclusion

---

## 7.1 Introduction

This thesis opened with a two-level research question. The main enquiry which motivated this project is how corpus stylistic methods can be usefully applied to the syntax of literary texts. To explore this problem a case study was needed and, therefore, I have used corpus stylistics to identify the differences between the syntax of an early Henry James novel and a late novel, which is written in his famous (or perhaps notorious) 'difficult' late style.

In this concluding chapter I will bring together the results of both parts of the project, beginning with Henry James's syntax and then discussing the use of corpus stylistics. The two novels which were used as examples were *Washington Square*, first published in 1880, and *The Golden Bowl*, published in 1904. Five chapters from each novel were compiled into a corpus, which was named the *Henry James Parsed Corpus (HJPC)*. The two parts of the corpus are referred to as *WS* and *GB*.

## 7.2 Henry James's syntax

### 7.2.1 Critical opinions of James's style

In Chapter 3 I reviewed some of the many critical works on James. As the James industry is vast, I was only able to pick out a few of the studies which are most relevant to my project. In particular I focused on discussions of James's complex and controversial late style.

James was a prolific analyst of his own and others' work, and his concerns are mirrored in the writings of others on his work. For example, James holds that a novel should reflect the writer's view of life and the experience of living, conveyed through the consciousness of a perceptive and intelligent character (who he calls a 'reflector') (James, 1948 [Original work published 1884]). Using the more modern term 'focalizer', this topic is discussed by Menikoff (1971) and Fowler (1993), although Chatman (1998) argues that the narrator is more present than James admits (see section 3.2.3). Many commentators (for example, Schwarz (1993), whose work is described in section 3.1) link James's style, and his use of focalizers, to the later, modernist development of a stream of consciousness style. Others, such as Welleck (1958) and Lubbock (1921) emphasise the influence of James's interest and experience in theatre. James himself describes the importance of

writing novels in scenes, which show rather than describe the action and contrast with pictorial episodes.

Opinions of James's complex late style are polarised. Blackmur (1935), Gifford (1983) and Leavis (1972 [Original work published 1948]) are all critical of what they see as overly ornate and impenetrable prose. On the other hand, Beebe (1968), Raleigh (1968), Mizener (1966) and Springer (1993) all approve of the way James writes his late novels, considering the complexity justified and necessary in pursuit of the subtle conveying of his protagonists' experience. A few commentators contrast the late style with an earlier and simpler version. For example, Raleigh specifically links the change in James's style with his increasing attempt to portray the consciousness of his characters. The same motivation, Raleigh suggests, explains the development of James's dramatic method of presentation. While Raleigh does not focus on language, he attempts to describe and explain differences between James's early, middle and late styles. Fowler takes a similar view, suggesting that the late novels differ from earlier ones in having fewer main protagonists but with greater focus on those characters and their mental lives. Cross (1993) engages more closely with James's syntax, developing a theory of compounding in the early novels which becomes more fragmented in the late works. This idea is partially supported by my figures – see Figure 5-5. The most thorough and quantitative analysis of James's early, middle and late prose is that of Chatman (1972). However, his focus is semantic rather than syntactic, as he is concerned to identify intangibility of reference.

My choice of James's syntax as a subject of study was motivated by the lack of work on the subject to date. There are few studies which focus specifically on James's literary language and their conclusions can be contradictory. Levin's analysis of the language of *The Ambassadors* surprisingly claims that it has 'relatively few subordinate clauses' (Levin, 1986, p. 27), though it is not clear what his comparator is. Levin also notes, as I do, James's frequent use of parenthesis. Springer, again writing about *The Ambassadors*, notices the variability in James's sentence structure which my analysis identifies. She also associates short, simple sentences towards the end of the book with the portrayal of strong emotion, a pattern which I observe in both *Washington Square* and *The Golden Bowl*.

Quantitative studies are even rarer and tend to use only short passages of prose. Watt's seminal paper (1960) is specifically only a description of the first paragraph of *The Ambassadors*. Short (1946) uses one chapter of the same novel but picks out the longer and more complex sentences for analysis. Within that sample, several characteristics of James's prose are identified which are echoed in my findings. These include a heavy use of parenthesis, the division of sentences into more than one free-standing unit, and a lack of differentiation of characters in

terms of their speech style. Leech and Short (2007) also identify James's heavy use of parenthesis, and note his increased use of dependent clauses in comparison with Conrad. However, this analysis is based only on James's short story 'The Pupil'. Hoover (2007) surveys the complete range of James's novels, finding evidence which groups them both chronologically and into three periods of writing. His evidence is purely lexical rather than syntactic.

### 7.2.2 Results of my analysis

While some of these critics have pointed out features of James's late style which echo my own findings, I have undertaken a quantitative corpus stylistic study which introduces an element of objectivity. I have explicitly compared one of James's early novels with a late novel in order to provide relevant comparisons between the features of his style. As it is a neglected area, I have specifically focused on syntax, and have compiled a parsed and annotated corpus of a sufficient size to enable statistically-significant results to be extracted which are applicable to the whole of each novel studied. Taking the critical theme of James's difficult late style, I have interpreted this difficulty as being partly due to syntactic complexity, and have contrasted that with coordination, which is a characteristic also identified by earlier critics. During the compilation of the *HJPC*, I noticed the prevalence of APU's (additional parsing units) and parenthesis but then I defined them rigorously so that they could be quantified. The results of this analysis were discussed in Chapter 5 and are summarised below.

The more complicated sentence structure of *The Golden Bowl* was suggested by the finding that *GB* had more, but not longer, clauses than *WS*. This was supported by the evidence that, while *WS* had more main than dependent clauses, the reverse was true of *GB*. Coordination had decreased from the early to the late novel.

Further investigation revealed that there were more speech clauses in *WS* than in *GB*, and that speech sentences included fewer dependent clauses. Separate analysis of clause types in speech and non-speech demonstrated that there is no increase in dependent clauses from the early to the late novel when James is writing dialogue. The ratio of main to dependent clauses in speech sentences is 1:1.5 in *WS* and 1:1.4 in *GB*. Far from there being a difficult late style in dialogue, James has remained remarkably consistent over the decades in this facet of his style.

In non-speech sentences, the style of *The Golden Bowl* can be characterised as more syntactically complex than that of *Washington Square*. In addition to this overall conclusion, analysis showed that *GB* has a small number of sentences which are extraordinarily complex, far exceeding *WS* in this respect. *GB* has a sentence

with 13 dependent clauses while no *WS* sentence has more than eight. However, most sentences in *GB* are comparable to those in *WS*.

The importance of using quantitative data was illustrated by the result of my analysis of the occurrence of APU's (additional parsing units). When compiling my corpus, it seemed to me that these were an important element of James's late style, echoing Short (1946). However, Figure 5-11 shows that there is no significant difference between the amount of APU's in *Washington Square* and *The Golden Bowl* with 3.89% and 3.78% of total clauses respectively. APU's have not been shown to be a distinguishing characteristic of James's late style.

Delay is a concept devised and defined syntactically for this project. It has been found to be a significant stylistic feature of *The Golden Bowl*. A similar process to that used for dependent clauses established that delay only slightly distinguishes the early and late styles in speech sentences (the ratio of instances of delay per word in *WS* compared to *GB* is 1:1.6). However, in non-speech there is a strong contrast. There are 2.19 instances of delay in *GB* for every instance in *WS*. This measure has proved to be a strong differentiator between the style of these early and late novels. However, as with dependency, the majority of sentences in *The Golden Bowl* have a low level of delay and are similar to *Washington Square's* sentences in this respect.

The question of the characteristics of James's late style is at the heart of this project. The results of the analysis of the *HJPC* cannot be definitely stated to represent the late style as a whole. Following Hoover (2007), I take that to be represented by the style of James last four novels as well as *The Ivory Tower*, which he left unfinished. Similarly, *Washington Square* is only one of the early novels which Hoover identifies (see Figure 3-1). However, my research is statistically significant for the whole of the two novels studied, and there is no reason to consider their style anomalous. In the process of this research project I have found that a very small number of sentences with extremely complex sentence structures make a disproportionate impact, contributing to readers' characterisation of James's late style as 'difficult'. This impact may seem to be a contradiction to the idea of foregrounding, which suggests that stylistically marked sentences are likely to be particularly significant in meaning at some level. It is this concept which informs my analysis in Chapter 6, reviewed in section 7.2.3. However, I believe there is no inherent contradiction here. If a reader is slowed by a deviant sentence and pays it more attention, the importance of that sentence may be noticed at that moment but an overall effect of working hard to decipher the text may linger. It may also be the case that in the late style, James goes too far in crafting exceptional sentences and some of them become almost impossible to decipher.

This dissertation does not describe all the syntactic differences between *Washington Square* and *The Golden Bowl*. Many other syntactic elements could be quantified and compared, and no doubt other resemblances and differences would be discovered. Moreover, lexis as well as syntax are contributors to James's late style. Hoover's (2007) analysis, which formed the basis for my text choices, shows that James's lexis changed chronologically. Other critics have highlighted semantic difficulties with James's late prose and the elaboration of his use of metaphor in his late novels. This dissertation has been able to identify some syntactic elements which do or do not distinguish James's syntax in these two novels, but this is not a complete characterisation of his style.

### 7.2.3 An alternative approach

A different use of the data contained in the *HJPC* was described in Chapter 6. A detailed corpus analysis of the syntax of the central chapter of each of my novels required new analytic formulae to be created. A coordination score was devised which included a count of all coordinated items in each sentence. In addition, a complexity score was formulated which brought together a wide number of different possible sentence features. Using these two measures and the length of sentences, three different profiles were compiled for each chapter. In each case, the profiles were found to mirror the plot and discourse types of the chapter and to facilitate the identification of the most foregrounded sentences by that measure. Using the profiles, the chapters could be divided into plot segments. For *WS18*, the segments mirrored the stages of Catherine's approach to her father and the development of their discussion. In *GB22*, the two main segments highlighted the division of the chapter into a description of the Prince's rumination and then of his conversation with Charlotte. Subsegments revealed subdivisions between those two main parts of the chapter.

Within dialogue episodes in each chapter, very emotional interchanges were related in short, simple sentences. Salient sentences, which were defined as those sentences which had outstandingly high scores in one or more of the word, coordination or complexity profiles, were found to convey significant moments in the plot, or aid an understanding of one of the protagonists. For example, *GB22:38* is a long sentence and the seventh most complex in the chapter. It signals the moment where the Prince realises that Charlotte is ready to move their relationship forward. None of the sentences around it have a similar complexity score (See section 6.6.1.4).

This method does not bring to light every important sentence or passage in these chapters, nor does it provide an automated literary criticism. However, it is a useful aid to a rigorous critical analysis, helping to provide a framework and key

points in a chapter which, in the case of GB22, is framed in complex and oblique language.

### 7.3 Corpus stylistics

My goal in applying corpus stylistics to literary texts was to use the evidence of quantitative data to guide and inform qualitative analysis. This echoes the interests of corpus stylisticians from the very beginning of this sub-discipline (and pre-computer quantitative literary critics). While this brings in an element of objectivity, complete objectivity is an impossible aim, and I would not even consider it a desirable one.

Corpus stylistics is now a maturing field with a number of different methodologies within the general description of using corpora to study literary texts (some would include other texts). Research projects can broadly be divided into those which use large natural language reference corpora and those in which a special corpus is compiled and annotated for the purpose. Large reference corpora can provide collocational data as a norm against which a literary work can be compared. An example of this approach is the work of Louw (1993) and others on semantic prosody. Semino and Short (2004) typify the alternative approach. They compiled their own corpus, annotating it by hand within using speech and thought presentation framework.

Various statistical methods have now been applied to literary texts as part of corpus stylistic research. Burrows has developed a number of statistical approaches for the analysis of texts, some of which he has used for authorship attribution and others for critical analysis. Biber et al. (1998) describe the application of multidimensional analysis in developing a functional analysis of texts. In their case, they use selections from existing large English language corpora.

Corpus stylistic studies can also be divided into those (the great majority) which analyse the lexis of texts and those which address syntax. Lexical research often uses the identification of collocates as an explanatory tool. Stubbs (2005) uses large corpora for collocational comparison with the work of Conrad and Culpeper (2002) uses a similar approach with Shakespearean texts. Both recognise that their research is led by the insights of previous critics. Mahlberg's (2007 and 2013) work on Dickens, and is close to this project in that it analyses the novels of a 19<sup>th</sup> century writer using corpus stylistics, although Mahlberg's corpus is very much larger than mine and her work focuses on lexis rather than syntax.

A very small number of studies use corpus stylistics to look at syntax. Hoover's (1999) analysis of *The Inheritors* includes a discussion of syntactic function. Mullender (2010) was restricted by having an unparsed corpus but



nevertheless described various different syntactic uses of *which* in Shakespeare's plays. Boyne (2009) explored the deviant syntax of *Riddley Walker* and *The Road* but was only able to analyse small sections of text. The small number and limited nature of these corpus stylistic studies of literary syntax illustrate the need for this research project.

Recent corpus stylistics studies, such as Mahlberg and Smith (2010) and Walker (2010) have combined a number of different methods in a literary analysis. Others have applied corpus stylistics to other approaches within stylistics in general.

While the *HJPC* has proved a powerful tool for analysing James's prose, the difficulties of hand parsing have limited its size. As far back as 1972, Milic was trying to devise an automated method of parsing but a complete and accurate automatic parser is still not available. Coh-Metrix (McNamara, et al., 2005) provides an online analysis tool but at the time of writing does not provide the quantitative data it has extracted to draw its conclusions, nor does it show its results on a sentence by sentence basis. An automatic analysis of the type ICECUP displays is still a distant prospect. Hand-parsing has advantages as insights into the texts develop during the process, as noted by Semino and Short (2004), but they also conclude that the onerous nature of the undertaking limits its use. In section 7.4.2 I discuss a possible alternative.

A further limitation of my research method could be considered to be its elements of subjectivity. The data in the *HJPC* can be objectively verified, and my parsing was not guided by any assumptions about Henry James's prose but by ICECUP conventions. However, the choice of which data to discuss, and which additional annotation to include, has a subjective aspect. I have been guided by a consensus of critical comment, and by a syntactic definition of complexity, but choices have still been made, and no doubt I have overlooked some possible alternative parameters. In Chapter 6, sentences to discuss were chosen because of their objectively exceptional qualities, but the cut off points were decided pragmatically, and the interpretation of the data was subjective. It was not my goal to remove subjectivity from corpus stylistics. Instead, I have worked with objectively-verifiable data to enable a rigorous analysis which can be replicated and evaluated by others.

## 7.4 Further research

### 7.4.1 Henry James's syntax

Further research on James's syntax would be desirable. There is a huge amount of data in the *HJPC* which has not been used in this study. It is hoped to make the

*HJPC* publicly available on the internet so that other researchers can use it, either to study James alone or to compare his texts with those of other writers. Figure 4-5 shows a few of the possibilities, as it includes unused annotation of tense, transitivity and dependent clause type, as well as pronoun and adverb type. These features would tie in with some of the suggestions made by James's critics. Watt (1960) discusses James's use of transitivity and negatives, and Lee (1923) and Norrman (1982) cite his unusual pronominal referencing. These ideas could be followed up using the *HJPC* and it would be possible to add further annotation if necessary.

The addition of some or all of the early and late novels to the *HJPC* would make it possible to test the hypothesis that my findings apply to James's early and late styles in general. Inclusion of novels from James's middle period of writing would allow research into the nature of their style, which might resemble the early or late versions or some intermediate stage, as is suggested by Hoover's lexical analysis (2007).

The corpus could also be extended to James's short stories and non-fiction writing. The latter would serve to clarify the question of whether James is deliberately writing in the late style for the purpose of his novels or whether that had become his idiolect towards the end of his life.

A more wide-ranging development would be to add texts by other authors. These might be contemporaries of James's or Modernists, whose writing is said by some to have been influenced by his style (c.f. Peter Wilson's (2004) comparison of *The Pickwick Papers* and *Ulysses* discussed in section 2.4).

#### **7.4.2 Corpus stylistics and syntax**

Two separate corpus stylistic methods were used in this dissertation, one using large quantities of text and the other relatively small amounts. Both yielded fruitful results in analysing Henry James's style. However, as discussed above, compiling a parsed corpus using ICECUP limited the amount of text which could be included because it was too time-consuming. Rather than parsing long sections of the novels, an alternative approach is being developed in which a sufficient quantity of sentences for statistical significance are chosen randomly, then parsed and analysed. This would facilitate further research of the kind exemplified in Chapter 5.

The method of short text analysis used in Chapter 6 can easily be extended. A piece of flash fiction ('That Colour' by John McGregor) has already been parsed as the first piece of a new 'Fiction' corpus, and salient sentences and other stylistic features were found to be helpful as part of a discussion of the story.

## 7.5 Conclusion

The two-level research question of this project has provided two levels of results. Henry James's syntax had not been examined in detail previously. It proved to be less homogeneous than is usually assumed, and the style of *The Golden Bowl* was found to be close to that of *Washington Square* in much of the text. However, many sentences in *The Golden Bowl* have exhibited great syntactic complexity and a confusing complexity of word order, which explains the difficulty of many readers in negotiating this novel.

Corpus stylistics has rarely been used to describe the syntax of literary texts. In this dissertation, the compilation of a parsed corpus and two methods for using that corpus for literary analysis have been described.

A number of new methods and measures have been devised for this project. This is the first time ICECUP has been used to compile a corpus from long literary texts. The concepts of delay and APU were formulated to describe James's style quantitatively. For detailed analysis of individual sentences, coordination and complexity scores were defined and added to a simple word count for each sentence. These new measures enabled the generation of chapter profiles, showing the distribution of each of the scores and revealing graphically the foregrounded salient sentences whose possible foregrounding effects could then be discussed.

Together with new insights about James's style, this dissertation presents a number of new corpus stylistic tools which make a significant contribution to the discipline.

# Bibliography

---

- Aarts, B., 1999. Corpus linguistics, Chomsky and Fuzzy Tree Fragments. *Language and Computers*, pp. 5-14.
- Aarts, B., 2008. *English Syntax and Argumentation*. 3rd ed. Basingstoke, Hampshire: Palgrave Macmillan.
- Andreev, V., 2008. A Multivariate Study of Style Differences in Poetry. In: G. Watson, ed. *The State of Stylistics*. Amsterdam, New York: Rodopi, pp. 117-190.
- Archer, D., 2009. Does Frequency Really Matter?. In: D. Archer, ed. *What's in a Word-list? Investigating Word Frequency and Keyword Extraction*. Farnham: Ashgate, pp. 1-15.
- Archer, D., Culpeper, J. & Rayson, P., 2009. Love - "a familiar or a devil"? An Exploration of Key Domains in Shakespeare's Comedies and Tragedies. In: D. Archer, ed. *What's in a Word-list?*. Farnham: Ashgate, pp. 137-157.
- Austen-Smith, B., 2004. The Counterfeit Symbol in Henry James's *The Golden Bowl*. *The Henry James Review*, 25(1), pp. 52-66.
- Bailey, R., 1969. Statistics and Style: a historical survey. In: L. Dolezel & R. Bailey, eds. *Statistics and Style*. New York: Elsevier, pp. 217-232.
- Bailey, R. W., 1989. The Future of Computational Stylistics. In: R. Potter, ed. *Literary Computing and Literary Criticism: Theoretical and Practical Essays on Theme and Rhetoric*. Philadelphia: University of Pennsylvania Press, pp. 3-12.
- Baker, P., Hardie, A. & McEnery, T., 2006. *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Barthes, R., 1971. Style and its image. In: S. Chatman, ed. *Literary Style: A Symposium*. Oxford: Oxford University Press, pp. 3-15.
- Beebe, M., 1968. The Turned Back of Henry James. In: T. Tanner, ed. *Henry James: Modern Judgements*. London: Macmillan, pp. 71-88.
- Bell, M., 1975. Style as Subject: Washington Square. *The Sewanee Review*, Winter, 83(1), pp. 19-38.
- Besant, W., 1884. *The Art of Fiction*. London: Chatto & Windus.
- Biber, D., 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber, D., 2011. Corpus linguistics and the study of literature: Back to the future?. *Scientific Study of Literature*, 1(1), pp. 15-23.
- Biber, D., Conrad, S. & Reppen, R., 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Birch, D., 1994. Stylistics. In: R. E. Asher, ed. *The Encyclopedia of Language and Linguistics*. Oxford: Pergamon Press, pp. 4378-4383.
- Blackmur, R., 1935. Introduction. In: R. Blackmur, ed. *The Art of the Novel: Critical Prefaces*. London: Charles Scribner's Sons, pp. vii-xxxix.

- Blackmur, R. P., 1959. Introduction. In: *Washington Square and The Europeans*. New York: Dell Publishing.
- Blakemore, D., 2009. Parentheticals and point of view in free indirect style. *Language and Literature*, May, 18(2), pp. 129-153.
- Boren, L. S., 1989. *Eurydice Reclaimed: Language, Gender and Voice in Henry James*. Ann Arbor, London: UMI Research Press.
- Boyne, M., 2009. *The Art of the (Deviant) Sentence: Syntactic Variation as a Stylistic Tool*. Middleburg, Online Proceedings of the Annual Conference of the Poetics and Linguistics Association (PALA).
- Bradbury, N., 2000. Introduction. In: *The Golden Bowl*. Ware, Hertfordshire: Wordsworth Editions Limited, pp. V-XXXII.
- Buitenhuis, P., 1970. *The Grasping Imagination*. Toronto and Buffalo: University of Toronto Press.
- Burnard, L., 2010. *British National Corpus*. [Online] Available at: [www.natcorp.ox.ac.uk/corpus/index.xml](http://www.natcorp.ox.ac.uk/corpus/index.xml) [Accessed 29 April 2013].
- Burrows, J., 1986. Modal Verbs and Moral Principles: An Aspect of Jane Austen's Style. *Literary and Linguistic Computing*, 1(1), pp. 9-23.
- Burrows, J., 1987. *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method*. Oxford: Clarendon Press.
- Burrows, J., 2002a. 'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship. *Literary and Linguistic Computing*, 17(3), pp. 267-287.
- Burrows, J., 2002b. The Englishing of Juvenal: Computational Stylistics and Translated Texts. *Style*, 36(4), pp. 677-679.
- Burrows, J., 2005. Andrew Marvell and the "Painter Satires": A Computational Approach to their Authorship. *The Modern Language Review*, April, 100(2), pp. 281-297.
- Burrows, J., 2007. All the Way Through: Testing for Authorship in Different Frequency Strata. *Literary and Linguistic Computing*, 22(1), pp. 27-47.
- Busse, B., 2011. Writing is medicine: blending cognitive and corpus stylistics. In: M. Callies, W. Keller & A. Lohöfer, eds. *Bi-Directionality in the Cognitive Sciences*. Amsterdam and Philadelphia: John Benjamins, pp. 121-156.
- Carroll, J. B., 1960. Vectors of Prose Style. In: *Style in Language*. Harvard, Mass.: John Wiley and Sons, pp. 283-292.
- Carter, R., 1994. Literary Language. In: R. E. Asher, ed. *The Encyclopedia of Language and Linguistics*. Oxford: Pergamon Press, pp. 2247-2249.
- Carter, R., 2003. Style. In: W. J. Frawley, ed. *International Encyclopedia of Linguistics*. Oxford: Oxford University Press, pp. 169-171.
- Carter, R., 2010. Methodologies for Stylistic Analysis: Practices and Pedagogies. In: D. McIntyre & B. Busse, eds. *Language and Style*. Basingstoke: Palgrave Macmillan, pp. 55-69.
- Carter, R. & Stockwell, P., 2008. Stylistics: retrospect and prospect. In: R. Carter & P. Stockwell, eds. *The Language and Literature Reader*. London and New York: Routledge, pp. 291-302.

- Chatman, S., 1971. Editor's Introduction. In: S. Chatman, ed. *Literary Style: A Symposium*. Oxford: Oxford University Press, pp. ix-xv.
- Chatman, S., 1972. *The Later Style of Henry James*. Oxford: Basil Blackwell.
- Chatman, S., 1998. Whatever happened to literary theory?. *The Southern Review*, 22 March, pp. 367-382.
- Cohn, D., 2001. "First Shock of Complete Perception": The Opening Episode of The Golden Bowl, Volume 2. *The Henry James Review*, 22(1), pp. 1-9.
- Conrad, S., 2000. Will Corpus Linguistics Revolutionize Grammar Teaching in the 21st Century?. *TESOL Quarterly*, 34(3), pp. 548-60.
- Cornell University Library, 2010. *Making of America: Harper's New Monthly Magazine*. [Online] Available at: <http://digital.library.cornell.edu/h/harp/index.html> [Accessed 24 January 2011].
- Craig, H., 1999. Contrast and Change in the Idiolects of Ben Jonson Characters. *Computers and the Humanities*, May, 33(3), pp. 221-240.
- Cross, M., 1993. *Henry James: The Contingencies of Style*. Basingstoke: MacMillan.
- Culpeper, J., 2002. Computers, language and characterisation: An Analysis of six characters in Romeo and Juliet. In: U. Melander-Marttala, C. Ostman & M. Kyto, eds. *Conversation in Life and Literature: Papers from the ASLA Symposium*. Uppsala: Association Suedoise de Linguistique Appliquee, pp. 11-30.
- Culpeper, J., 2009. Keyness: words, parts-of-speech and semantic categories in the character-talk of Shakespeare's Romeo and Juliet. *International Journal of Corpus Linguistics*, 14(1), pp. 29-59.
- Cureton, R. D., 2003. Linguistics and Literature: Stylistics. In: W. J. Frawley, ed. *International Encyclopedia of Linguistics*. Oxford: Oxford University Press, pp. 466-472.
- da Costa Fialho, O., 2007. Foregrounding and refamiliarization: understanding readers' response to literary texts. *Language and Literature*, 16(2), pp. 105-123.
- Dryden, E., 2010. The Imp of the Perverse: Metaphor in The Golden Bowl. *The Henry James Review*, 31(2), pp. 111-124.
- Edel, L., 1955. *The Selected Letters of Henry James*. New York: Farrar, Straus and Cudhay.
- Espinal, M. T., 1991. The Representation of Disjunct Constituents. *Language*, 67(4), pp. 726-762.
- Facchinetti, R., 2007. *Corpus Linguistics 25 years on*. Amsterdam and New York : Rodopi.
- Firth, J., 1957. A synopsis of linguistic theory, 1930-1955. *Studies in Linguistic Analysis*, Issue Special Volume of the Philological Society, pp. 1-32.
- Fish, S., 1980. What is Stylistics and Why Are They Saying Such Terrible Things About It?. In: S. Fish, ed. *Is There a Text in This Class?: The Authority of Interpretive Communities*. Cambridge, Massachusetts: Harvard University Press, pp. 68-96.

- Fletcher, W. H., 2012. *kfNgram*. [Online] Available at: <http://www.kwicfinder.com/kfNgram/kfNgramHelp.html> [Accessed 16 April 2013].
- Fludernik, M., 1993. *The Fictions of Language and the Languages of Fiction*. London: Routledge.
- Fowler, R., 1981. *Literature as Social Discourse: The Practice of Linguistic Criticism*. London: Batsford Academic and Educational.
- Fowler, R., 1996. *Linguistic Criticism*. 2nd ed. Oxford: Oxford University Press.
- Fowler, V. C., 1993. The Later Fiction. In: *A Companion to Henry James Studies*. Westport, Connecticut and London: Greenwood Press, pp. 179-205.
- Frantz Blackall, J., 1993. The Experimental Period. In: D. M. Fogel, ed. *A Companion to Henry James Studies*. Westport, Connecticut: Greenwood Press, pp. 147-178.
- Fucks, W., 1952. On Mathematical Analysis of Style. *Biometrika*, Volume 39, pp. 122-129.
- Gale, R. L., 1964. *The Caught Image: Figurative language in the fiction of Henry James*. Chapel Hill: University of North Carolina Press.
- Garvin, P. L., 1964. *A Prague School Readers on Esthetics, Literary Structure, and Style*. Washington, D.C.: Georgetown University Press.
- Geisler, C., 2002. Investigating register variation in nineteenth-century English: a multi-dimensional comparison. In: R. Reppen, S. Fitzmaurice & D. Biber, eds. *Using Corpora to Explore Linguistic Variation*. Amsterdam/Philadelphia: John Benjamins, pp. 249-271.
- Gifford, H., 1983. Henry James: The Drama of Discrimination. In: B. Ford, ed. *The New Pelican Guide to English Literature*. London: Penguin, pp. 115-130.
- Givón, T., 2009. *The Genesis of Syntactic Complexity: Diachrony, ontogeny, neuro-cognition, evolution*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Graesser, A. C., Dowell, N. & Moldovan, C., n.d. *A Computer's Understanding of Literature*. [Online] Available at: <http://tinyurl.com/l2594d2> [Accessed 1 October 2013].
- Graesser, A. C., McNamara, D. S., Louwerse, M. M. & Cai, Z., 2004. Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments & Computers*, 36(2), pp. 193-202.
- Greenbaum, S. & Nelson, G., 1995. Clause relationships in spoken and written English. *Functions of Language*, 2(1), pp. 1-21.
- Hallen, C. L. & Shakespear, J., 2002. The T-Unit as a Measure of Syntactic Complexity in Emily Dickinson's Poems. *The Emily Dickinson Journal*, 11(1), pp. 91-103.
- Halliday, M., 1971. Linguistic function and literary style: an inquiry into the language of William Golding's "The Inheritors". In: S. Chatman, ed. *Literary Style: A Symposium*. Oxford: Oxford University Press, pp. 330-365.
- Haralson, E., 2006. Henry James. In: D. S. Kastan, ed. *The Oxford Encyclopedia of British Literature*. Oxford: Oxford University Press, pp. 127-133.

- Hardy, D., 2004. Collocational analysis as a stylistic discovery procedure: the case of Flannery O'Connor's Eyes. *Style*, Winter, 38(4), pp. 92-116.
- Harper Collins, n.d. *About the Collins Corpus and the Bank of English*. [Online] Available at: [www.mycobuild.com/about-collins-corpus.aspx](http://www.mycobuild.com/about-collins-corpus.aspx) [Accessed 3 December 2013].
- Hart, M., 2010. *Project Gutenberg*. [Online] Available at: [www.gutenberg.org](http://www.gutenberg.org) [Accessed 24 January 2011].
- Hathaway, R. D., 2010. *The Henry James scholar's Guide to Web Sites*. [Online] Available at: <http://www2.newpaltz.edu/~hathaway/> [Accessed 24 January 2011].
- Heywood, J., Semino, E. & Short, M., 2002. Linguistic metaphor identification in two extracts from novels. *Language and Literature*, February, 11(1), pp. 35-54.
- Hocks, R. A., 1993. From Literary Analysis to Postmodern Theory: A Historical Narrative of James Criticism. In: D. M. Fogel, ed. *A Companion to Henry James Studies*. Westport, Connecticut and London: Greenwood Press, pp. 3-24.
- Hoover, D., 1999. *Language and Style in 'The Inheritors'*. Lanham, Maryland: University Press of America.
- Hoover, D., 2007. Corpus Stylistics, Stylometry, and the Styles of Henry James. *Style*, Summer, 41(2), pp. 174-203.
- Hoover, D., 2008. Searching for style in Modern American Poetry. In: S. Zyngier, M. Borolussi, A. Chesnokova & J. Auracher, eds. *Directions in Empirical Literary Studies*. Amsterdam/Philadelphia: John Benjamins, pp. 212-227.
- Hopkins, V., 1968. Visual Art Devices and Parallels in James. In: T. Tanner, ed. *Henry James: Modern Judgements*. London: Macmillan, pp. 89-127.
- Hori, M., 2004. *Investigating Dickens' Style*. Basingstoke: Palgrave Macmillan.
- Horne, P., 1997. The Master and the 'Queer Affair' of 'The Pupil'. In: N. H. Reeve, ed. *Henry James: The Shorter Fiction*. Basingstoke: Macmillan, pp. 114-137.
- Horne, P., 2005. Henry James among the Poets. *The Henry James Review*, 26(1), pp. 68-81.
- Ho, Y., 2011. *Corpus Stylistics in Principles and Practice*. London and New York : Continuum.
- Huddleston, R., 1984. *Introduction to the Grammar of English*. Cambridge: Cambridge University Press.
- Huddleston, R. & Pullum, G. K., 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Hunston, S., 2007. Semantic Prosody Revisited. *International Journal of Corpus Linguistics*, 12(2), pp. 249-268.
- Jakobson, R., 1960. Closing statement: Linguistics and Poetics. In: T. A. Sebeok, ed. *Style in Language*. Cambridge, Mass.: The Technology Press of Massachusetts Institute of Technology, pp. 350-377.



- James, H., 1935. *The Art of the Novel: Critical Prefaces*. London: Charles Scribner's Sons.
- James, H., 1948 [Original work published 1884]. The Art of Fiction. In: M. Roberts, ed. *The Art of Fiction and Other Essays*. New York: Oxford University Press, pp. 3-23.
- James, H., 1998 [Original work published 1880]. *Washington Square*. Oxford: Oxford University Press.
- James, H., 1998 [Original work published 1903]. *The Ambassadors*. Oxford and New York: Oxford University Press.
- James, H., 2000 [Original work published 1904]. *The Golden Bowl*. Ware, Hertfordshire: Wordsworth Editions Limited.
- Jeffries, L., 2010. 'The Unprofessionals': Syntactic Iconicity and Reader Interpretation in Contemporary Poems. In: D. McIntyre & B. Busse, eds. *Language and Style*. Basingstoke: Palgrave Macmillan, pp. 95-115.
- Jeffries, L. & McIntyre, D., 2010. *Stylistics*. Cambridge: Cambridge University Press.
- Johnson, R. G., 1971. *A study of the style of HJ's Late Novels*. Ohio: Bowling Green State University.
- Kenny, A., 1992. *Computers and the Humanities*. London: The British Library.
- Knights, L. C., 1975. Henry James and Human Liberty. *The Sewanee Review*, Winter, 83(1), pp. 1-18.
- Korte, B., 1997. *Body language in literature*. Toronto and London: University of Toronto Press.
- Laitinen, T., 1975. *Aspects of Henry James's Style*. Helsinki: Suomalainen Tiedeakatemia.
- Lambert, M., 1981. *Dickens and the Suspended Quotation*. New Haven and London: Yale University Press.
- Le Fanu, M., 1982. Introduction. In: *Washington Square*. Oxford: Oxford University Press, pp. vii-xxvii.
- Leavis, F. R., 1972 [Original work published 1948]. *The Great Tradition*. Harmondsworth, Middlesex: Penguin.
- Lebowitz, N., 1968. Magic and Metamorphosis in The Golden Bowl. In: T. Tanner, ed. *Henry James: Modern Judgements*. London : Macmillan, pp. 327-339.
- Leech, G., 1969. *A Linguistic Guide to English Poetry*. Harlow, Essex: Longman.
- Leech, G., 2008. *Language in Literature: Style and Foregrounding*. Harlow, Essex: Pearson.
- Leech, G. & Short, M., 1981. *Style in Fiction*. London and New York: Longman.
- Leech, G. & Short, M., 2007. *Style in Fiction: A Linguistic Introduction to English Fictional Prose*. 2nd ed. Harlow: Pearson Longman.
- Lee, V., 1992 [Original work published 1923]. *The Handling of Words*. Lewiston, Queenston, Lampeter: Edwin Mellen Press.
- Levin, H., 1986. Introduction. In: *The Ambassadors*. London: Penguin, pp. 7-29.

- Levin, S. R., 1965. Internal and External Deviation in Poetry. *Word*, 21(2), pp. 225-237.
- Llewellyn Smith, V., 2009. Introduction. In: *The Golden Bowl*. Oxford: Oxford University Press, pp. vii-xxxi.
- Lodge, D., 1966. *The Language of Fiction*. London and New York: Routledge and Kegan Paul.
- Lodge, D., 2002. *The Language of Fiction*. 3rd ed. London and New York: Routledge and Kegan Paul.
- Louw, B., 1993. Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies. In: *Text and Technology*. Philadelphia/Amsterdam: John Benjamins, pp. 157-176.
- Louw, B., 1997. The role of corpora in critical literary appreciation. In: A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles, eds. *Teaching and Language Corpora*. London and New York: Longman, pp. 240-251.
- Lubbock, P., 1921. *The Craft of Fiction*. London: Jonathan Cape.
- Lucas, J., 1972. Washington Square. In: J. Goode, ed. *The Air of Reality: New Essays on Henry James*. London: Methuen, pp. 36-59.
- Mahlberg, M., 2007. Corpus stylistics: bridging the gap between linguistic and literary studies. In: M. Hoey, M. Mahlberg, M. Stubbs & W. Teubert, eds. *Text, Discourse and Corpora*. London, New York: Continuum, pp. 219-246.
- Mahlberg, M., 2013. *Corpus Stylistics and Dickens's Fiction*. New York and London: Routledge.
- Mahlberg, M. & Smith, C., 2010. Corpus Approaches to Prose Fiction: Civility and Body Language in *Pride and Prejudice*. In: D. McIntyre & B. Busse, eds. *Language and Style*. Basingstoke: Palgrave Macmillan, pp. 449-467.
- Martindale, C., 2007. Deformation forms the course of literary history. *Language and Literature*, 16(2), pp. 141-153.
- Maseychik, W. J., 1968. Points of Departure from *The American*. In: T. Tanner, ed. *Henry James: Modern Judgements*. London: Macmillan, pp. 116-127.
- McEnery, T. & Hardie, A., 2012. *Corpus Linguistics*. Cambridge: Cambridge University Press.
- McIntyre, D., 2010. Dialogue and Characterization in *Reservoir Dogs*. In: D. McIntyre & B. Busse, eds. *Language and Style*. Basingstoke: Palgrave Macmillan, pp. 162-182.
- McNamara, D. S., Louw, B., Cai, Z. & Graesser, A., 2005. *Coh-Metrix version 1.4*. [Online] Available at: <http://cohmetrix.memphis.edu> [Accessed 25 December 2013].
- Mendenhall, T. C., 1901. A Mechanical Solution of a Literary Problem. *Popular Science Monthly*, December, pp. 97-105.
- Menikoff, B., 1971. The Subjective Pronoun in the late style of Henry James. *English Studies*, October, 52(5), pp. 436-441.
- Miles, J., 1971. Style as Style. In: *Literary style: a Symposium*. Oxford: Oxford University Press, pp. 24-28.

- Milic, L. T., 1967. A Quantitative approach to the style of Jonathan Swift. The Hague, Paris: Mouton & Co.
- Milic, L. T., 1972. Autocoding in Computational Stylistics. In: B. B. Kachru & H. F. W. Stahlke, eds. *Current Trends in Stylistics*. Edmonton, Alberta: Linguistic Research Inc., pp. 263-273.
- Mizener, A., 1966. Introduction. In: S. G. Putt, ed. *A Reader's Guide to Henry James*. London: Thames and Hudson, pp. 7-24.
- Mullender, J., 2010. Shakespeare's late syntax: a comparative analysis of which relativisation in the dramatic works. Unpublished doctoral thesis: University of Birmingham.
- Nelson, G., Wallis, S. & Aarts, B., 2002a. International Corpus of English Corpus Utility Program 3.1, s.l.: s.n.
- Nelson, G., Wallis, S. & Aarts, B., 2002b. Exploring Natural Language: Working with the British Component of the International Corpus of English. Amsterdam and Philadelphia: John Benjamins.
- Norrman, R., 1982. *The Insecure World of Henry James's Fiction*. London and Basingstoke: Macmillan.
- O'Halloran, K., 2007a. Corpus-assisted literary evaluation. *Corpora*, 2(1), pp. 33-63.
- O'Halloran, K., 2007b. The subconscious in James Joyce's 'Eveline': a corpus stylistic analysis that chews on the 'Fish hook'. *Language and Literature*, 16(3), pp. 227-244.
- Ohmann, R., 1964. Generative Grammars and the Concept of Literary Style. *Word: Journal of the Linguistic Circle of New York*, 20(3), pp. 423-439.
- Osgood, C., 1960. Some effects of motivation on style of encoding. In: T. A. Sebeok, ed. *Style in Language*. Cambridge, Mass.: Technology Press of Massachusetts Institute of Technology, pp. 293-306.
- Oxford English Dictionary, 1991. *The Compact Oxford English Dictionary*. 2nd ed. Oxford: Clarendon Press.
- Philip, G., 2011. Colouring Meaning: Collocation and connotation in figurative language. Amsterdam and Philadelphia: John Benjamins.
- Poirier, R., 1984. Melodrama, Irony and the Comic Method. In: A. Shelston, ed. *Washington Square and The Portrait of a Lady*. London: Macmillan, pp. 46-59.
- Poole, A., 2010. Introduction. In: *Washington Square*. Oxford: Oxford University Press, pp. vii-xxvii.
- Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J., 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Raleigh, J. H., 1968. Henry James: The Poetics of Empiricism. In: T. Tanner, ed. *Henry James: Modern Judgements*. London: Macmillan, pp. 52-70.
- Rayson, P., 2009. *Wmatrix: a web-based corpus processing environment*. [Online] Available at: <http://ucrel.lancs.ac.uk/wmatrix/> [Accessed 15 January 2014].

- Renouf, A., 2007. Corpus development 25 years on: from super-corpus to cyber-corpus. In: R. Facchinetti, ed. *Corpus Linguistics 25 years on*. Amsterdam and New York: Rodopi, pp. 27-49.
- Rudman, J., 2006. Authorship Attribution: Statistical and Computational Methods. In: K. Brown, ed. *Encyclopedia of Language and Linguistics*. 2nd ed. Oxford: Elsevier, pp. 611-617.
- Schwarz, D. R., 1993. James's Theory of Fiction and Its Legacy. In: D. M. Fogel, ed. *A Companion to Henry James Studies*. Westport, Connecticut: Greenwood Press, pp. 39-53.
- Scott, M., 2004. *Wordsmith Tools version 4*. Oxford: Oxford University Press.
- Sebeok, T. A., 1960. Introduction. In: *Style in Language*. Cambridge, Mass.: Technology Press of Massachusetts Institute of Technology, pp. 1-5.
- Semino, E. & Culpeper, J., 2002. Foreword. In: E. Semino & J. Culpeper, eds. *Cognitive Stylistics: Language and Cognition in text analysis*. Amsterdam and Philadelphia: John Benjamins, pp. IX-XVI.
- Semino, E. & Short, M., 2004. Corpus Stylistics: speech, writing and thought presentation in a corpus of English writing. London: Routledge.
- Shelston, A., 1984. *Washington Square and The Portrait of a Lady: A Casebook*. London: Macmillan.
- Short, M., 1996. *Exploring the Language of Poems, Plays and Prose*. London and New York: Longman.
- Short, R. W., 1946. The Sentence Structure of Henry James. *American Literature*, May, 18(2), pp. 71-88.
- Sinclair, J., 1966. Beginning the Study of Lexis. In: C. Bazell, J. C. Catford, M. Halliday & R. Robins, eds. *In Memory of J. R. Firth*. London: Longmans, pp. 410-430.
- Sinclair, J., 1987. The Nature of the Evidence. In: J. Sinclair, ed. *Looking Up: An account of the COBUILD Project in lexical computing*. London, Glasgow: Collins, pp. 150-159.
- Smit, D. W., 1988. *The Language of a Master: Theories of Style and the Late Writing of Henry James*. Carbondale and Edwardsville : Southern Illinois University Press.
- Sopčák, P., 2007. "Creation from nothing": a foregrounding study of James Joyce's drafts for *Ulysses*. *Language and Literature*, 16(2), pp. 183-196.
- Springer, M. D., 1993. Closure in James: A Formalist Feminist View. In: D. M. Fogel, ed. *A Companion to Henry James Studies*. Westport, Connecticut: Greenwood Press, pp. 265-283.
- Stevenson, B., 1989. Adapting Hypothesis Testing to a Literary Problem. In: *Literary Computing and Literary Criticism: Theoretical and Practical Essays on Theme and Rhetoric*. Philadelphia: University of Pennsylvania Press, pp. 61-72.
- Stewart, D., 2010. *Semantic Prosody: a critical evaluation*. London: Routledge.
- Stubbs, M., 2001. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.

- Stubbs, M., 2005. Conrad in the computer: examples of quantitative stylistic methods. *Language and Literature*, 14(1), pp. 5-24.
- Svartvik, J., 2007. Corpus linguistics 25+ years on. In: *Corpus Linguistics 25 Years on*. Amsterdam and New York: Rodopi, pp. 11-25.
- Szmrecsanyi, B., 2004. *On operationalizing syntactic complexity*. Louvain-la-Neuve, Belgium, s.n., pp. 1032-1039.
- Tanner, T., 1968. Introduction. In: T. Tanner, ed. *Henry James: Modern Judgements*. London: Macmillan, pp. 11-41.
- Taylor, A., 2003. Henry James. In: S. R. Serafin & V. Grosvenor Myer, eds. *The Continuum Encyclopedia of British Literature*. New York, London: Continuum, pp. 517-521.
- Tennyson, L. A., 1983 [Original work published in various versions from 1842-1899]. *Idylls of the King*. Harmondsworth: Penguin.
- Toolan, M., 1990. *The Stylistics of Fiction*. London: Routledge.
- Toolan, M., 2009. Narrative Progression in the Short Story: A corpus stylistic approach. Amsterdam/Philadelphia: John Benjamins.
- Traugott, E. C. & Pratt, M. L., 2008. Language, linguistics and literary analysis. In: R. Carter & P. Stockwell, eds. *The Language and Literature Reader*. London and New York: Routledge, pp. 34-38.
- van Peer, W., Jakemuldeer, J. & Zyngier, S., 2007. Lines on feeling: foregrounding, aesthetics and meaning. *Language and Literature*, 16(2), pp. 197-213.
- Verdonk, P., 2006. Style. In: K. Brown, ed. *Encyclopedia of Language and Linguistics*. Oxford: Elsevier, pp. 196-210.
- Wales, K., 2006. Stylistics. In: K. Brown, ed. *Encyclopedia of Language and Linguistics*. Oxford: Elsevier, pp. 213-217.
- Wales, K., 2011. *A Dictionary of Stylistics*. 3rd ed. Harlow: Pearson Education.
- Walker, B., 2010. Wmatrix, Key Concepts and the Narrators in Julian Barnes's Talking It Over. In: D. McIntyre & B. Busse, eds. *Language and Style*. Basingstoke: Palgrave Macmillan, pp. 364-387.
- Wallis, S. A., 2008. *Maketag 1.0*. [Online] Available at: [www.ucl.ac.uk/english-usage/staff/sean/resources/maketag.zip](http://www.ucl.ac.uk/english-usage/staff/sean/resources/maketag.zip) [Accessed 6 June 2013].
- Watt, I., 1960. The First Paragraph of The Ambassadors: An Explication. *Essays in Criticism*, 10(3), pp. 250-74.
- Welleck, R., 1958. Henry James's Literary Theory and Criticism. *American Literature*, November, 30(3), pp. 293-321.
- Widdowson, H., 2008. On the Deviance of Literary Discourse. In: R. Carter & P. Stockwell, eds. *The Language and Literature Reader*. London and New York: Routledge, pp. 294-306.
- Williams, C., 1970. *Style and Vocabulary*. London: Griffin.
- Wilson, A., 1968. Charles Dickens: A Haunting. In: A. Dyson, ed. *Dickens: Modern Judgements*. London: Macmillan, pp. 30-39.

- Wilson, P., 2004. The Corpus of Jinglese: A Syntactic Profile of an Idiolectal 'System of Stenography'. *Critical Survey*, 16(3), pp. 78-93.
- Wilson, R., 1981. *Henry James's Ultimate Narrative: The Golden Bowl*. St. Lucia, London, New York: University of Queensland Press.
- Wimsatt, W. K., 1941. *The Prose Style of Samuel Johnson*. New Haven: Yale University Press.
- Wright, W., 1968. Maggie Verver: Neither Saint nor Witch. In: T. Tanner, ed. *Henry James: Modern Judgements*. London: Macmillan, pp. 316-326.
- Wynne, M., 2006. Stylistics: Corpus Approaches. In: K. Brown, ed. *Encyclopedia of Language and Linguistics*. 2nd ed. Oxford: Elsevier, pp. 223-226.
- Wynne, M., 2012. Review of Corpus Stylistics in Principle and Practice by Yufang Ho. *Literary and Linguistic Computing*, 27(4), pp. 474-476.
- Yeazell, R. B., 1976. *Language and Knowledge in the Late Novels of Henry James*. Chicago and London: University of Chicago Press.

# Appendix 1

## *Washington Square Chapter 18*

---

This text is reproduced, with sentence numbers added, from the e-text used to create the first part of the *HJPC*. The e-text was created from the novel published by Macmillan in London in 1881. There are a few minor differences in punctuation between the two texts; these are marked as footnotes. Sentences found to be salient in the three word profiles discussed in Chapter 6 are shown in bold (see section 6.5.2).

1. Catherine sat alone by the parlour fire - sat there for more than an hour, lost in her meditations.
2. **Her aunt seemed to her aggressive and foolish, and to see it so clearly - to judge Mrs. Penniman so positively – made her feel old and grave.**
3. She did not resent the imputation of weakness; it made no impression on her, for she had not the sense of weakness, and she was not hurt at not being appreciated.
4. **She had an immense respect for her father, and she felt that to displease him would be a misdemeanour analogous to an act of profanity in a great temple:<sup>56</sup> but her purpose had slowly ripened, and she believed that her prayers had purified it of its violence.**
5. The evening advanced, and the lamp burned dim without her noticing it; her eyes were fixed upon her terrible plan.
6. She knew her father was in his study - that he had been there all the evening; from time to time she expected to hear him move.
7. She thought he would perhaps come, as he sometimes came, into the parlour.
8. At last the clock struck eleven, and the house was wrapped in silence; the servants had gone to bed.
9. Catherine got up and went slowly to the door of the library, where she waited a moment, motionless.
10. Then she knocked, and then she waited again.
11. Her father had answered her, but she had not the courage to turn the latch.
12. What she had said to her aunt was true enough - she was afraid of him; and in saying that she had no sense of weakness she meant that she was not afraid of herself.
13. She heard him move within, and he came and opened the door for her.
14. "What is the matter?" asked the Doctor.
15. "You are standing there like a ghost."
16. She went into the room, but it was some time before she contrived to say what she had come to say.
17. **Her father, who was in his dressing-gown and slippers, had been busy at his writing-table, and after looking at her for some moments, and**

---

<sup>56</sup> Digital text has semi-colon. Corrected to colon to match both *Cornhill Magazine* August 1880 and Macmillan edition 1881

**waiting for her to speak, he went and seated himself at his papers again.**

18. His back was turned to her - she began to hear the scratching of his pen.
19. **She remained near the door, with her heart thumping beneath her bodice; and she was very glad that his back was turned, for it seemed to her that she could more easily address herself to this portion of his person than to his face.**
20. At last she began, watching it while she spoke.
21. **"You told me that if I should have anything more to say about Mr. Townsend you would be glad to listen to it."**
22. "Exactly, my dear," said the Doctor, not turning round, but stopping his pen.
23. Catherine wished it would go on, but she herself continued.
24. "I thought I would tell you that I have not seen him again, but that I should like to do so."
25. "To bid him good-bye?" asked the Doctor.
26. The girl hesitated a moment.
27. "He is not going away."
28. The Doctor wheeled slowly round in his chair, with a smile that seemed to accuse her of an epigram; but extremes meet, and Catherine had not intended one.
29. "It is not to bid him good-bye, then?" her father said.
30. "No, father, not that; at least,<sup>57</sup> not for ever.
31. I have not seen him again, but I should like to see him," Catherine repeated.
32. The Doctor slowly rubbed his under lip<sup>58</sup> with the feather of his quill.
33. "Have you written to him?"
34. "Yes, four times."
35. "You have not dismissed him, then.
36. Once would have done that."
37. "No," said Catherine; "I have asked him - asked him to wait."
38. Her father sat looking at her, and she was afraid he was going to break out into wrath; his eyes were so fine and cold.
39. "You are a dear, faithful child," he said<sup>59</sup> at last.
40. "Come here to your father."
41. And he got up, holding out his hands toward her.
42. The words were a surprise, and they gave her an exquisite joy.
43. She went to him, and he put his arm round her tenderly, soothingly; and then he kissed her.
44. After this he said - "Do you wish to make me very happy?"
45. "I should like to - but I am afraid I can't," Catherine answered.
46. "You can if you will.
47. It all depends on your will."
48. "Is it to give him up?" said Catherine.
49. "Yes, it is to give him up."
50. And he held her still, with the same tenderness, looking into her face and resting his eyes on her averted eyes.
51. There was a long silence; she wished he would release her.
52. "You are happier than I, father," she said, at last.

---

<sup>57</sup> Comma not present in *Cornhill Magazine*, present in Macmillan edition

<sup>58</sup> *Under-lip* in *Cornhill Magazine*, in this form in Macmillan edition

<sup>59</sup> Digital text has comma here. Deleted to match both *Cornhill Magazine* and Macmillan edition



53. "I have no doubt you are unhappy just now.
54. But it is better to be unhappy for three months and get over it,<sup>60</sup> than for many years and never get over it."
55. "Yes, if that were so," said Catherine.
56. "It would be so; I am sure of that."
57. She answered nothing, and he went on.<sup>61</sup>
58. "Have you no faith in my wisdom, in my tenderness, in my solicitude for your future?"
59. "Oh, father!" murmured the girl.
60. "Don't you suppose that I know something of men: their vices, their follies, their falsities?"
61. She detached herself, and turned upon him.
62. "He is not vicious - he is not false!"
63. Her father kept looking at her with his sharp, pure eye.
64. "You make nothing of my judgment, then?"
65. "I can't believe that!"
66. "I don't ask you to believe it, but to take it on trust."
67. Catherine was far from saying to herself that this was an ingenious sophism; but she met the appeal none the less squarely.
68. "What has he done - what do you know?"
69. "He has never done anything - he is a selfish idler."
70. "Oh, father, don't abuse him!" she exclaimed, pleadingly.
71. "I don't mean to abuse him; it would be a great mistake.
72. You may do as you choose," he added, turning away.
73. "I may see him again?"
74. "Just as you choose."
75. "Will you forgive me?"
76. "By no means."
77. "It will only be for once."
78. "I don't know what you mean by once.
79. You must either give him up or continue the acquaintance."
80. "I wish to explain - to tell him to wait."
81. "To wait for what?"
82. "Till you know him better - till you consent."
83. "Don't tell him any such nonsense as that.
84. I know him well enough, and I shall never consent."
85. "But we can wait a long time," said poor Catherine, in a tone which was meant to express the humblest conciliation, but which had upon her father's nerves the effect of an iteration not characterised by tact.
86. The Doctor answered, however, quietly enough: "Of course you can wait till I die, if you like."
87. Catherine gave a cry of natural horror.
88. "Your engagement will have one delightful effect upon you; it will make you extremely impatient for that event."
89. Catherine stood staring, and the Doctor enjoyed the point he had made.
- 90. It came to Catherine with the force - or rather with the vague impressiveness - of a logical axiom which it was not in her province to**

<sup>60</sup> No comma here in digital text but added to match both *Cornhill Magazine* and Macmillan edition

<sup>61</sup> Colon in *Cornhill Magazine* becomes a full stop in Macmillan edition and digital text, dividing sentences 57 and 58

**controvert; and yet, though it was a scientific truth, she felt wholly unable to accept it.**

91. "I would rather not marry, if that were true," she said.
- 92. "Give me a proof of it, then; for it is beyond a question that by engaging yourself to Morris Townsend you simply wait for my death."**
93. She turned away, feeling sick and faint; and the Doctor went on.<sup>62</sup>
94. "And if you wait for it with impatience, judge, if you please, what HIS eagerness will be!"
95. Catherine turned it over - her father's words had such an authority for her that her very thoughts were capable of obeying him.
96. There was a dreadful ugliness in it, which seemed to glare at her through the interposing medium of her own feebler reason.
97. Suddenly, however, she had an inspiration - she almost knew it to be an inspiration.
98. "If I don't marry before your death, I will not after," she said.
- 99. To her father, it must be admitted, this seemed only another epigram; and as obstinacy, in unaccomplished minds, does not usually select such a mode of expression, he was the more surprised at this wanton play of a fixed idea.**
100. "Do you mean that for an impertinence?" he inquired; an inquiry of which, as he made it, he quite perceived the grossness.
101. "An impertinence?"
102. Oh, father, what terrible things you say!"
103. "If you don't wait for my death, you might as well marry immediately; there is nothing else to wait for."
104. For some time Catherine made no answer; but finally she said - "I think Morris - little by little - might persuade you."
105. "I shall never let him speak to me again.
106. I dislike him too much."
- 107. Catherine gave a long, low sigh; she tried to stifle it, for she had made up her mind that it was wrong to make a parade of her trouble, and to endeavour to act upon her father by the meretricious aid of emotion.**
- 108. Indeed, she even thought it wrong - in the sense of being inconsiderate - to attempt to act upon his feelings at all; her part was to effect some gentle, gradual change in his intellectual perception of poor Morris 's character.**
109. But the means of effecting such a change were at present shrouded in mystery, and she felt miserably helpless and hopeless.
110. She had exhausted all arguments, all replies.
111. Her father might have pitied her, and in fact he did so; but he was sure he was right.
- 112. "There is one thing you can tell Mr. Townsend, when you see him again," he said:<sup>63</sup> "that if you marry without my consent, I don't leave you a farthing of money.**
113. That will interest him more than anything else you can tell him."

---

<sup>62</sup> Colon in *Cornhill Magazine* becomes a full stop in Macmillan edition and digital text, dividing sentences 93 and 94

<sup>63</sup> Comma here in digital text but changed to colon to match both *Cornhill Magazine* and Macmillan edition

114. "That would be very right," Catherine answered.
115. "I ought not in that case to have a farthing of your money."
116. "My dear child, "the Doctor observed, laughing, "your simplicity is touching.
117. Make that remark, in that tone, and with that expression of countenance, to Mr. Townsend, and take a note of his answer.
- 118. It won't be polite - it will express irritation; and I shall be glad of that, as it will put me in the right; unless, indeed - which is perfectly possible - you should like him the better for being rude to you."**
119. "He will never be rude to me," said Catherine, gently.
120. "Tell him what I say, all the same. "
121. She looked at her father, and her quiet eyes filled with tears.
122. "I think I will see him, then, "she murmured, in her timid voice.
123. "Exactly as you choose!"
124. And he went to the door and opened it for her to go out.
125. The movement gave her a terrible sense of his turning her off.
126. "It will be only once, for the present," she added, lingering a moment.
127. "Exactly as you choose," he repeated, standing there with his hand on the door.
128. "I have told you what I think.
129. If you see him, you will be an ungrateful, cruel child; you will have given your old father the greatest pain of his life."
130. This was more than the poor girl could bear; her tears overflowed, and she moved towards her grimly consistent parent with a pitiful cry.
131. Her hands were raised in supplication, but he sternly evaded this appeal.
132. Instead of letting her sob out her misery on his shoulder, he simply took her by the arm and directed her course across the threshold, closing the door gently but firmly behind her.
133. After he had done so, he remained listening.
134. For a long time there was no sound; he knew that she was standing outside.
135. He was sorry for her, as I have said; but he was so sure he was right.
136. At last he heard her move away, and then her footstep creaked faintly upon the stairs.
137. The Doctor took several turns round his study, with his hands in his pockets, and a thin sparkle, possibly of irritation, but partly also of something like humour, in his eye.
138. "By Jove," he said to himself, "I believe she will stick - I believe she will stick!"
139. And this idea of Catherine "sticking" appeared to have a comical side, and to offer a prospect of entertainment.
140. He determined, as he said to himself, to see it out.

## Appendix 2

### *The Golden Bowl Chapter 22*

---

This text is reproduced, with sentence numbers added, from the e-text used to create the second part of the *HJPC*. Sentences were found to be salient in the three word profiles discussed in Chapter 6 are shown in bold (see section 6.6.2).

1. **It was quite for the Prince after this as if the view had further cleared; so that the half-hour during which he strolled on the terrace and smoked - the day being lovely - overflowed with the plenitude of its particular quality.**
2. **Its general brightness was composed doubtless of many elements, but what shone out of it as if the whole place and time had been a great picture, from the hand of genius, presented to him as a prime ornament for his collection and all varnished and framed to hang up - what marked it especially for the highest appreciation was his extraordinarily unchallenged, his absolutely appointed and enhanced possession of it.**
3. **Poor Fanny Assingham's challenge amounted to nothing: one of the things he thought of while he leaned on the old marble balustrade - so like others that he knew in still more nobly-terraced Italy - was that she was squared, all-conveniently even to herself, and that, rumbling toward London with this contentment, she had become an image irrelevant to the scene.**
4. **It further passed across him - as his imagination was, for reasons, during the time, unprecedentedly active - that he had after all gained more from women than he had ever lost by them; there appeared so, more and more, on those mystic books that are kept, in connexion with such commerce, even by men of the loosest business habits, a balance in his favour that he could pretty well as a rule take for granted.**
5. **What were they doing at this very moment, wonderful creatures, but trying to outdo each other in his interest? - from Maggie herself, most wonderful in her way of all, to his hostess of the present hour, into whose head it had so inevitably come to keep Charlotte on, for particular reasons, and who had asked in this benevolent spirit why in the world, if not obliged, without plausibility, to hurry, her husband's son-in-law shouldn't wait over in her company.**
6. He would at least see, Lady Castledean had said, that nothing dreadful should happen to her either while still there or during the exposure of the run to town; and, for that matter, if they exceeded a little their licence it would positively help them to have done so together.
7. Each of them would in this way have the other comfortably to complain of at home.
8. **All of which, besides, in Lady Castledean as in Maggie, in Fanny Assingham as in Charlotte herself, was working for him without provocation or pressure, by the mere play of some vague sense on**

**their part - definite and conscious at the most only in Charlotte - that he wasn't, as a nature, as a character, as a gentleman, in fine, below his remarkable fortune.**

9. But there were more things before him than even these; things that melted together, almost indistinguishably, to feed his sense of beauty.
10. **If the outlook was in every way spacious - and the towers of three cathedrals, in different counties, as had been pointed out to him, gleamed discernibly, like dim silver, in the rich sameness of tone - didn't he somehow the more feel it so because, precisely, Lady Castledean had kept over a man of her own, and that this offered a certain sweet intelligibility as the note of the day?**
11. It made everything fit; above all it diverted him to the extent of keeping up, while he lingered and waited his meditative smile.
12. She had detained Charlotte because she wished to detain Mr. Blint, and she couldn't detain Mr. Blint, disposed though he clearly was to oblige her, without spreading over the act some ampler drapery.
13. **Castledean had gone up to London; the place was all her own; she had had a fancy for a quiet morning with Mr. Blint, a sleek civil accomplished young man - distinctly younger than her ladyship - who played and sang delightfully (played even "bridge" and sang the English-comic as well as the French-tragic), and the presence - which really meant the absence - of a couple of other friends, if they were happily chosen, would make everything all right.**
14. **The Prince had the sense, all good-humouredly, of being happily chosen, and it wasn't spoiled for him even by another sense that followed in its train and with which during his life in England he had more than once had reflectively to deal: the state of being reminded how after all, as an outsider, a foreigner, and even as a mere representative husband and son-in-law, he was so irrelevant to the working of affairs that he could be bent on occasion to uses comparatively trivial.**
15. No other of her guests would have been thus convenient for their hostess; affairs, of whatever sorts, had claimed, by early trains, every active easy smoothly-working man, each in his way a lubricated item of the great social political administrative *engrenage* - claimed most of all Castledean himself, who was so very oddly, given the personage and the type, rather a large item.
16. If he, the great and the clever Roman, on the other hand, had an affair, it wasn't of that order; it was of the order verily that he had been reduced to as to a not quite glorious substitute.
17. It marked however the feeling of the hour with the Prince that this vision of being 'reduced' interfered not at all with the measure of his actual ease.
18. **It kept before him again at moments the so familiar fact of his sacrifices - down to the idea of the very relinquishment, for his wife's convenience, of his real situation in the world; with the consequence thus that he was, in the last analysis, among all these so often inferior people, practically held cheap and made light of.**
19. **But though all this was sensible enough there was a spirit in him that could rise above it, a spirit that positively played with the facts, with all of them; from that of the droll ambiguity of English relations to**

**that of his having in mind something quite beautiful and independent and harmonious, something wholly his own.**

20. He couldn't somehow take Mr. Blint seriously - he was much more an outsider, by the larger scale, even than a Roman prince who consented to be in abeyance.
21. Yet it was past finding out, either, how such a woman as Lady Castledean could take him - since this question but sank for him again into the fathomless depths of English equivocation.
22. **He knew them all, as was said, "well"; he had lived with them, stayed with them, dined, hunted, shot and done various other things with them; but the number of questions about them he couldn't have answered had much rather grown than shrunken, so that experience struck him for the most part as having left in him but one residual impression.**
23. They didn't like *les situations nettes* - that was all he was very sure of.
24. They wouldn't have them at any price; it had been their national genius and their national success to avoid them at every point.
25. **They called it themselves, with complacency, their wonderful spirit of compromise - the very influence of which actually so hung about him here from moment to moment that the earth and the air, the light and the colour, the fields and the hills and the sky, the blue-green counties and the cold cathedrals, owed to it every accent of their tone.**
26. Verily, as one had to feel in presence of such a picture, it had succeeded; it had made, up to now, for that seated solidity in the rich sea-mist on which the garish, the supposedly envious, peoples have ever cooled their eyes.
27. But it was at the same time precisely why even much initiation left one at given moments so puzzled as to the element of staleness in all the freshness and of freshness in all the staleness, of innocence in the guilt and of guilt in the innocence.
28. There were other marble terraces, sweeping more purple prospects, on which he would have known what to think, and would have enjoyed thereby at least the small intellectual fillip of a discerned relation between a given appearance and a taken meaning.
29. **The enquiring mind, in these present conditions, might, it was true, be more sharply challenged; but the result of its attention and its ingenuity, it had unluckily learned to know, was too often to be confronted with a mere dead wall, a lapse of logic, a confirmed bewilderment.**
30. And moreover above all nothing mattered, in the relation of the enclosing scene to his own consciousness, but its very most direct bearings.
31. Lady Castledean's dream of Mr. Blint for the morning was doubtless already, with all the spacious harmonies re-established, taking the form of 'going over' something with him, at the piano, in one of the numerous smaller rooms that were consecrated to the less gregarious uses; what she had wished had been effected - her convenience had been assured.
32. This made him however ask himself the more where Charlotte was - since he didn't at all suppose her to be making a tactless third, which would be to have accepted mere spectatorship, in the duet of their companions.

33. The upshot of everything for him, alike of the less and of the more, was that the exquisite day bloomed there like a large fragrant flower that he had only to gather.
34. But it was to Charlotte he wished to make the offering, and as he moved along the terrace, which rendered visible parts of two sides of the house, he looked up at all the windows that were open to the April morning and wondered which of them would represent his friend's room.
35. It befell thus that his question was after no long time answered; he saw Charlotte appear above as if she had been called by the pausing of his feet on the flags.
36. She had come to the sill, on which she leaned to look down, and she remained there a minute smiling at him.
37. He had been immediately struck with her wearing a hat and a jacket - which conduced to her appearance of readiness not so much to join him, with a beautiful uncovered head and a parasol, where he stood, as to take with him some larger step altogether.
- 38. The larger step had been since the evening before intensely in his own mind, though he hadn't fully thought out even yet the slightly difficult detail of it; but he had had no chance, such as he needed, to speak the definite word to her, and the face she now showed affected him thereby as a notice that she had wonderfully guessed it for herself.**
39. They had these identities of impulse - they had had them repeatedly before; and if such unarranged but unerring encounters gave the measure of the degree in which people were, in the common phrase, meant for each other, no union in the world had ever been more sweetened with rightness.
40. What in fact most often happened was that her rightness went, as who should say, even further than his own; they were conscious of the same necessity at the same moment, only it was she who as a general thing most clearly saw her way to it.
41. Something in her long look at him now out of the old grey window, something in the very poise of her hat, the colour of her necktie, the prolonged stillness of her smile, touched into sudden light for him all the wealth of the fact that he could count on her.
42. He had his hand there, to pluck it, on the open bloom of the day; but what did the bright minute mean but that her answering hand was already intelligently out?
43. So therefore while the minute lasted it passed between them that their cup was full; which cup their very eyes, holding it fast, carried and steadied and began, as they tasted it, to praise.
44. He broke however after a moment the silence.
45. "It only wants a moon, a mandolin and a little danger to be a serenade."
46. "Ah then," she lightly called down, "let it at least have THIS!"
47. With which she detached a rich white rosebud from its company with another in the front of her dress and flung it down to him.
48. He caught it in its fall, fixing her again after she had watched him place it in his buttonhole.
49. "Come down quickly!" he said in an Italian not loud but deep.
50. "*Vengo, vengo!*" she as clearly, but more lightly, tossed out; and she had left him the next minute to wait for her.

51. He came along the terrace again, with pauses during which his eyes rested, as they had already often done, on the brave darker wash of far-away water-colour that represented the most distant of the cathedral towns.
52. **This place, with its great church and its high accessibility, its towers that distinguishably signalled, its English history, its appealing type, its acknowledged interest, this place had sounded its name to him half the night through, and its name had become but another name, the pronounceable and convenient one, for that supreme sense of things which now throbbed within him.**
53. He had kept saying to himself "Glo'ster, Glo'ster, Glo'ster," quite as if the sharpest meaning of all the years just ended were intensely expressed in it.
54. That meaning was really that his situation remained quite sublimely consistent with itself, and that they absolutely, he and Charlotte, stood there together in the very lustre of this truth.
55. Every present circumstance helped to proclaim it; it was blown into their faces as by the lips of the morning.
56. **He knew why he had from the first of his marriage tried with such patience for such conformity; he knew why he had given up so much and bored himself so much; he knew why he had at any rate gone in, on the basis of all forms, on the basis of his having in a manner sold himself, for a *situation nette*.**
57. It had all been just in order that his - well, what on earth should he call it but his freedom? - should at present be as perfect and rounded and lustrous as some huge precious pearl.
58. He hadn't struggled nor snatched; he was taking but what had been given him; the pearl dropped itself, with its exquisite quality and rarity, straight into his hand.
59. Here precisely it was, incarnate; its size and its value grew as Mrs. Verver appeared, afar off, in one of the smaller doorways.
60. She came toward in silence while he moved to meet her; the great scale of this particular front, at Matcham, multiplied thus, in the golden morning, the stages of their meeting and the successions of their consciousness.
61. It wasn't till she had come quite close that he produced for her his "Glo'ster, Glo'ster, Glo'ster," and his "Look at it over there! "
62. She knew just where to look.
63. "Yes - isn't it one of the best?"
64. There are cloisters or towers or something."
65. And her eyes, which, though her lips smiled, were almost grave with their depths of acceptance, came back to him.
66. "Or the tomb of some old king."
67. "We must see the old king; we must 'do' the cathedral," he said; "we must know all about it.
68. If we could but take," he exhaled, "the full opportunity!"
69. And then while, for all they seemed to give him, he sounded again her eyes: "I feel the day like a great gold cup that we must somehow drain together."
70. "I feel it, as you always make me feel everything, just as you do; so that I know ten miles off how you feel!
71. But do you remember," she asked, "apropos of great gold cups, the beautiful one, the real one, that I offered you so long ago and that you wouldn't have?
72. Just before your marriage" - she brought it back to him: "the gilded crystal bowl in the little Bloomsbury shop."



73. "Oh yes!" - but it took, with a slight surprise on the Prince's part, some small recollecting.
74. "The treacherous cracked thing you wanted to palm off on me, and the little swindling Jew who understood Italian and who backed you up!
75. But I feel this an occasion," he immediately added, "and I hope you don't mean," he smiled, "that AS an occasion it's also cracked."
76. They spoke, naturally, more low than loud, overlooked as they were, though at a respectful distance, by tiers of windows; but it made each find in the other's voice a taste as of something slowly and deeply absorbed.
77. "Don't you think too much of 'cracks' and aren't you too afraid of them?
78. I risk the cracks," said Charlotte, "and I've often recalled the bowl and the little swindling Jew, wondering if they've parted company.
79. He made," she said, "a great impression on me."
80. "Well, you also, no doubt, made a great impression on him, and I dare say that if you were to go back to him you'd find he has been keeping that treasure for you.
81. But as to cracks," the Prince went on - "what did you tell me the other day you prettily call them in English? 'rifts within the lute'? - risk them as much as you like for yourself, but don't risk them for me."
82. He spoke it in all the gaiety of his just barely-tremulous serenity.
83. "I go, as you know, by my superstitions.
84. And that's why," he said, "I know where we are.
85. They're every one to-day on our side."
86. Resting on the parapet toward the great view she was silent a little, and he saw the next moment that her eyes were closed.
87. "I go but by one thing."
88. Her hand was on the sun-warmed stone; so that, turned as they were away from the house, he put his own upon it and covered it.
89. "I go by YOU," she said.
90. "I go by you."
91. So they remained a moment, till he spoke again with a gesture that matched.
92. "What's really our great necessity, you know, is to go by my watch.
93. It's already eleven" - he had looked at the time; "so that if we stop here to luncheon what becomes of our afternoon?"
94. To this Charlotte's eyes opened straight.
95. "There's not the slightest need of our stopping here to luncheon.
96. Don't you see," she asked, "how I'm ready?"
97. He had taken it in, but there was always more and more of her.
98. "You mean you've arranged - ?"
99. "It's easy to arrange.
100. My maid goes up with my things.
101. You've only to speak to your man about yours, and they can go together."
102. "You mean we can leave at once?"
103. She let him have it all.
104. "One of the carriages, about which I spoke, will already have come back for us.
105. If your superstitions are on our side," she smiled, "so my arrangements are, and I'll back my support against yours."
106. "Then you had thought," he wondered, "about Gloucester?"
107. She hesitated - but it was only her way.
108. "I thought YOU would think.

109. We have, thank goodness, these harmonies.
110. They're food for superstition if you like.
111. It's beautiful," she went on, "that it should be Gloucester; 'Glo'ster Glo'ster,' as you say, making it sound like an old song.
- 112. However, I'm sure 'Glo'ster Glo'ster' will be charming," she still added; "we shall be able easily to lunch there, and, with our luggage and our servants off our hands, we shall have at least three or four hours.**
113. We can wire," she wound up, "from there."
114. Ever so quietly she had brought it, as she had thought it, all out, and it had to be as covertly that he let his appreciation expand.
115. "Then Lady Castledean - ?"
116. "Doesn't dream of our staying."
117. He took it, but thinking yet.
118. "Then what does she dream - ?"
119. "Of Mr. Blint, poor dear; of Mr. Blint only."
120. Her smile for him - for the Prince himself - was free.
121. "Have I positively to tell you that she doesn't want us?"
122. She only wanted us for the others - to show she wasn't left alone with him.
123. Now that that's done and that they've all gone she of course knows for herself - !"
124. "Knows'?" the Prince vaguely echoed.
125. "Why that we like cathedrals; that we inevitably stop to see them, or go round to take them in, whenever we've a chance; that it's what our respective families quite expect of us and would be disappointed for us to fail of.
126. This, as *forestieri*," Mrs. Verver pursued, "would be our pull - if our pull weren't indeed so great all round."
127. He could only keep his eyes on her.
128. "And have you made out the very train - ?"
129. "The very one.
130. Paddington - the 6.50 'in'.
131. That gives us oceans; we can dine, at the usual hour, at home; and as Maggie will of course be in Eaton Square I hereby invite you."
132. For a while he still but looked at her; it was a minute before he spoke.
133. "Thank you very much.
134. With pleasure."
135. To which he in a moment added: "But the train for Gloucester?"
136. "A local one - 11:22; with several stops, but doing it a good deal, I forget how much, within the hour.
137. So that we've time.
138. "Only," she said, "we must employ our time."
139. He roused himself as from the mere momentary spell of her; he looked again at his watch while they moved back to the door through which she had advanced.
140. But he had also again questions and stops - all as for the mystery and the charm.
141. "You looked it up - without my having asked you?"
142. "Ah my dear," she laughed, "I've seen you with Bradshaw!
143. It takes Anglo-Saxon blood."
144. "Blood'?" he echoed.
145. "You've that of every race!"

146. It kept her before him.
147. "You're terrible."
148. Well, he could put it as he liked.
149. "I know the name of the inn."
150. "What is it then?"
151. "There are two - you'll see.
152. But I've chosen the right one.
153. And I think I remember the tomb," she smiled.
154. "Oh the tomb - !"
155. Any tomb would do for him.
156. "But I mean I had been keeping my idea so cleverly for you while there you already were with it."
157. "You had been keeping it 'for' me as much as you like.
158. But how do you make out," she asked, "that you were keeping it FROM me?"
159. "I don't - now.
160. How shall I ever keep anything - some day when I shall wish to?"
161. "Ah for things I mayn't want to know I promise you shall find me stupid."
162. They had reached their door, where she herself paused to explain.
163. "These days, yesterday, last night, this morning, I've wanted everything."
164. Well, it was all right.
165. "You shall HAVE everything."