Narrative Progression in the Short Story
A corpus stylistic approach

Michael Toolan

John Benjamins Publishing Company
Narrative Progression in the Short Story
Linguistic Approaches to Literature (LAL)

Linguistic Approaches to Literature (LAL) provides an international forum for researchers who believe that the application of linguistic methods leads to a deeper and more far-reaching understanding of many aspects of literature. The emphasis will be on pragmatic approaches intersecting with areas such as discourse analysis, sociolinguistics, ethnolinguistics, rhetoric, philosophy, cognitive linguistics, psycholinguistics and stylistics.

Editors

Willie van Peer
University of Munich

Sonia Zyngier
Federal University of Rio de Janeiro

Advisory Editorial Board

Timothy R. Austin
Loyola University Chicago

Douglas Biber
Northern Arizona University

Lubomir Dolezel
University of Toronto

Donald C. Freeman
University of Southern California

Harald Fricke
University of Fribourg

Raymond W. Jr. Gibbs
University of California, Santa Cruz

Rachel Giora
Tel Aviv University

Paisley Livingston
University of Copenhagen

Colin Martindale
University of Maine

Sara Mills
Sheffield Hallam University

Mick Short
Lancaster University

Michael Toolan
University of Birmingham

Reuven Tsur
Tel Aviv University

Jean Jacques Weber
University Centre Luxemburg

Peter Verdonk
University of Amsterdam

Volume 6

Narrative Progression in the Short Story. A corpus stylistic approach
by Michael Toolan
Narrative Progression in the Short Story
A corpus stylistic approach

Michael Toolan
University of Birmingham

John Benjamins Publishing Company
Amsterdam/Philadelphia
# Table of contents

Acknowledgements ........................................... ix
List of figures and tables ...................................... xi

CHAPTER 1
Introduction: Narrative prospecting .......................... 1
1. The crucial question: how does text ‘guide’ the reader? 1
2. A brief sketch of the research context and methods .... 3
3. On genre, relevance, scripts, and background ......... 5
4. Guided expectation ............................................. 8
5. Text integrity .................................................... 10
6. Predictive reading .............................................. 11
7. Conclusion ..................................................... 12

CHAPTER 2
Collocation and corpus stylistics ............................ 15
1. Studies in discourse prospection and expectation ..... 15
2. Collocational stylistics ....................................... 18
3. Choosing an appropriate comparator corpus ......... 25
4. A note on the stories selected for analysis ............ 28
5. Conclusion ..................................................... 29

CHAPTER 3
Lexical patternings in short stories ......................... 31
1 Word frequencies in “Two Gallants” ................. 31
   1.1 “Gallantry” in *Dubliners* ......................... 31
   1.2 High and low frequency lexis in “Two Gallants” 33
2. Textual richness measured by lexical diversity ...... 34
   2.1 The type-token measure ............................... 34
   2.2 Lexical innovation and text structure ............ 37
3. Foregrounding via repetitive phrasing or by novelty of phrasing and collocation 40
   3.1 Clusters ..................................................... 40
3.2 Achieving the stern task of living: a tissue of not-so-vague associations 42

4. Keyword analysis of “Two Gallants” 47
4.1 Using keywords on “Two Gallants”: section by section 50

CHAPTER 4
Top keyword sentences as story waymarking 53
1. A top keyword abridgement of “Two Gallants” 53
2. Applying the top keyword procedure in story abridgements 59
   2.1 Updike’s “A & P” 59
   2.2 Carver’s “Boxes” 63
   2.3 Carver’s “Cathedral” 66
3. Further procedural questions about the top keyword method 68
   3.1 Distinguishing the top keyword from other frequent keywords. 68
   3.2 On the importance, in the top keyword, of proper name status. 70
4. Some interim conclusions 71

CHAPTER 5
Keywords and the language of guidance in “The Love of a Good Woman” 77
1. Story opening as initial guidance 77
2. Top lexical keywords as narrativity indices 87
3. Textual segmentation and keywords’ collocate 90
4. Local (within-section) interrelation and collocation of keywords 93

CHAPTER 6
Repetition and para-repetition in story structure 97
1. A more delicate keywords and plotlinks analysis of “The Love of a Good Woman” 97
2. Keyword personal pronouns and idiolect-signalling 100
3. The non-repetitive echo: long-distance patterning via associated lexis and analogy 103
   3.1 Rubbing and scratching surfaces 103
   3.2 Dark above, light below 104
   3.3 Sorrowful plummeting 105
   3.4 Bashing, banging and braining 106
4. Para-repetitive narrative bonding between story opening and remainder 108
### Table of contents

**CHAPTER 7**  
Prospection and expectation: Core signalling 113  
1. Sentences featuring named main characters 114  
   1.1 High frequency and keyword character referencing 114  
   1.2 Modelling cohesive chains in long texts by sampling 116  
2. Narrative-tense finite verbs in character-depicting action clauses 119  
   2.1 Narrative-tense verbs with inquits excluded 120  
   2.2 VVD density 123  
   2.3 Is independent capture of narrative-tense action verbs needed? 125  
3. The cueing power of first sentences of narrative paragraphs 126  
4. Narrativity carried by “fully lexical” frequent keywords and clusters 128  
   4.1 Frequent keywords 128  
   4.2 Clusters 132  

**CHAPTER 8**  
Prospection and expectation: Embedded signalling 135  
1. The heightened narrativity of characters’ represented thought 136  
   1.1 Automating identification of FIT:  
      rule 1, narrative modal verbs with pronouns 137  
   1.2 Exclude modals following if, whether, that… 138  
   1.3 Automating identification of FIT:  
      Rule 2, include all questions and exclamations in the narrative 139  
   1.4 Automating identification of FIT:  
      Rule 3, include all flanking sentences containing subjective modals 140  
   1.5 Modifying the three FIT-finding procedures 140  
2. Prospective direct speech 145  
3. Negation-carrying clauses 148  
4. Narrative verbs of modality and mental processing 150  
5. Implementing the model with “Two Gallants” 152  
6. Compiling the abridgement 157  

**CHAPTER 9**  
The textual tracking of suspense and surprise 165  
1. Narrative suspense 165  
2. Narrative surprise 171  
3. Textualising suspense in “Two Gallants” 174  
4. Textualising surprise in “Bliss” and “A Small, Good Thing” 179
CHAPTER 10
Next steps 189
1. Corpus-based study of narrativity: a work in progress. 189
2. Expanding and refining the model: modality and evaluation 192
3. The reader’s experience of the text 195
4. Directions for future research 196

References 201

Name index 209

Topic index 211
Acknowledgements

Many friends and colleagues have assisted me in many ways, direct or indirect, in the writing of this book, including being the source of ideas and inspiration. I particularly wish to thank the following: Michael Barlow, Beatrix Busse, Ronald Carter, Carmen Rosa Caldas-Coulthard, Malcolm Coulthard, Pernilla Danielsson, George Dillon, Catherine Emmott, José Ángel García Landa, Roy Harris, Michael Hoey, Susan Hunston, Susan Lohafer, Norman Macleod, Miguel Ángel Martínez Cabeza, Oliver Mason, Rosamund Moon, Nina Nørgaard, Charles Owen, Ruth Page, John Pier, Paul Rayson, Mike Scott, Mick Short, the much-missed John Sinclair, and Michael Stubbs.

Kees Vaes was my admirably efficient editor at Benjamins, and Willie van Peer and Sonia Zyngier edit the series in which this volume appears; I am most grateful to all three for their support and welcome, after I had encountered a little publishing ‘situation’ elsewhere. Both series editors took a personal interest in the book, and I found this hugely encouraging. I want to thank Sonia and Willie for the meticulous care with which they read and suggested corrections and improvements to not one but two full drafts of the book. Such editorial diligence and engagement is, in my experience, truly exceptional.

Preliminary versions of parts of these chapters were first presented at a variety of conferences: the conference of the International Association for Literary Semantics in Krakow; conferences of the European Society for the Study of English, in Strasbourg and Zaragoza; conferences on the modern short story at Salamanca and Lisbon; a Narratology Workshop at the University of Southern Denmark, Kolding; and the annual conference of the Society for the Study of Narrative Literature, Washington D.C. I was especially fortunate to be invited to run a week-long workshop, in July 2007, on corpus linguistic approaches to narrative progression in stories, for a lively and suitably questioning group of scholars. The workshop formed part of a hugely enjoyable International Systemic-Functional Linguistics Congress Summer School, at the University of Southern Denmark, Odense. I have benefited from colleagues’ comments and questions following all these presentations.

A particular friend, guide and critic, to whom I am especially indebted, is Wolfgang Teubert. He was extraordinarily generous with his time and expertise,
making numerous valuable comments on draft chapters. As for my debts to Julianne Statham, these cannot be reduced to words.

Early versions of parts of Chapters Two, Three, Seven and Eight have appeared in the following items listed in the Bibliography: Toolan 2004a, b, and c; 2006; 2008a and b.

The early research for and writing of this book were made possible by a study leave from the University of Birmingham; completion of the project was enabled by a Research Leave Award from the Arts and Humanities Research Council (UK). I am most grateful for both these sources of support.
List of figures and tables

Table 1. Keywords in “Two Gallants” 49
Table 2. Keywords in Updike’s “A & P” 59
Table 3. Positive lexical keyword listing for “The Love of a Good Woman” 88
Figure 1. Keywords by story section (frequency of 4 or more occurrences in the section) 91
Figure 2. A keywords map for “The love of a good woman” 92
Table 4. Positive keywords, subsection by subsection, ≥ 3 instances 98
Table 5. Trends in pronoun use in “Bliss” 116
Figure 3. Collocational trends before Street in “Two gallants” 129
Figure 4. Summary of the eight proposed parameters (and search procedures) of narrative prospection 164
CHAPTER 1

Introduction

Narrative prospecting

1. The crucial question: how does text ‘guide’ the reader?

This study begins with the assumption that readers of narratives are ‘guided’, in their experience and expectations concerning narrative progression, by the text’s precise wording. But we currently have only a limited account of how narrative wording guides expectations and responses; the present study adds to the account from a corpus analytic and literary linguistic orientation. It aims to contribute to a fuller understanding of how, in written narrative, material that is either explicit or at least implicit in the text gives rise to such distinct impressionistic reader judgements as ones of suspense, surprise, secrecy or gaps, mystery, tension, obscurity, and even incoherence.

These judgements are often regarded by narratologists as core ingredients of ‘narrativity’: the sense of narrative dynamism or progression that a reader ascribes to the text. This study develops and defends the idea that there are distinct kinds of lexical and phrasal construction and patterning in a written narrative, dispersed through the text, that particularly contribute to that narrativity and the reader’s reactions such as suspense or surprise. Representative 20th century short stories (by James Joyce, Katherine Mansfield, Raymond Carver, Alice Munro, and others) are the ‘test-bed’ for all the procedures and commentary that follow, so it must be emphasised that the effects of narrative progression identified here may be distinctive of just the modern short story genre. To explore the textualised narrativity of these stories, I use literary linguistic and more particularly corpus analytic methods and tools, believing these can help to reveal and identify the expectation-guiding textual patterns I am interested in.

The broad goal of the book, then, is to contribute to a better understanding of the role that lexico-phrasal patterning in narrative sentences plays in the shaping of readers’ expectations as to narrative progression: what in the text has passed, is passing, and is to pass. For the claims and descriptions to have full credibility, however, they must eventually be tested against actual readers’ responses – as best as we can examine them. This is always a problematic issue, for many reasons. One is the ‘observer’s paradox’, by which probing a situation to determine its elements
affects the situation to the point that some of those elements are altered. Another is the possibility of ‘contamination’ of a reader’s verbal comment or evaluation by the conditions of its elicitation, cultural expectations, or peer-group pressure. And having been elicited, is it real? Is it the reader’s actual response to the text, or his/her response to the researcher’s question (about the text)? Whatever the words or behaviour are that accompany or follow readers’ readings of a poem, can we be sure that that is their response, rather than merely epiphenomenal to their response (an introductory treatment for Humanities scholars is van Peer, W., S. Zyn- gier, and F. Hakemulder, eds. (2007).

Readers’ judgements of the dynamism and interest in a narrative can be expressed in varied ways and elicited by a series of procedures. But often, readers will comment on how they felt that a particular place in the narrative delivered a surprise, or that the narration was unforthcoming to the point of being secretive, or slow, or confusing, or seemingly unconnected with previous text. In the longer run, comprehensive studies of narrative progression need to link text-based evidence of the kind discussed here with the testing and questioning of readers. The latter topics are important aspects of the larger project, but cannot be treated at length in the present book: there is too much ground to cover first with regard to identifying and describing the textual resources for generating expectations in modern short stories. A detailed empirical study of readers’ responses to stories, which may serve to confirm the validity of the expectation-structuring and patterning proposed here, must be the subject of later work. But that there is a causal link between the signal- lings of narrative progression in a story and numberless readers’ individual first-person experiences of suspense, or acceleration, or secrecy is a crucial conviction. That conviction holds even if the particular textual bases of progression that are proposed here are judged questionable or in need of revision.

The research questions identified above inevitably give rise to subsidiary ones, and this chapter addresses some of these. For example, how relevant is the specificity of a text’s lexis (lexical choice) to narrative and narrativity? In Forster’s famous example – The king died, the queen died of grief (Forster 1927: 87) – how important is the first use of died and would alternative lexis carry equal narrativity or not, and why (e.g., The king expired. Then the queen died of grief)? And how useful are corpus and computational methods when the sequential aspect of a text is the focus of interest, as it is here? Many corpus analytic methods are designed to help lay bare the patterns in a language or genre as a single stable situation, a repertoire with numerous identified gambits and rare, odd or unacceptable combinations of form and use. Can some corpus methods be adapted to address the rather different business of textual progress and processing, particularly if the genre of text under scrutiny is narrative, where a change or development in the situation can be ex- pected and is in fact criterial?
Chapter 1. Introduction

2. A brief sketch of the research context and methods

How literary narratives proceed and how our expectations in the course of them are created by the text (and subsequently are met, or thwarted) is an ancient topic. But it has attracted fresh treatments in recent years from cognitivist psychologists and psycholinguists (Mandler 1984, Bruner 1986, van Dijk and Kintsch 1983, Gerrig 1993, Gibbs 1994, Graesser et al. 1994 and 1997, Schank 1995). Their proposals have only rarely been taken up and applied by literary linguists, who might have been expected to try to match the psychological theories (e.g., theories of story processing) with detailed analyses of the verbal material of literary texts. Stylistic studies of longer narrative texts, even where oriented to textual sequence, have tended to look at single phenomena – metaphors, themes, motifs, modes of discourse representation, anaphora and naming – rather than at texture as a whole, and its progressive modulation or change (one prominent exception has been the invaluable work on 'mind style': see Fowler 1986, Leech and Short 1981, Werth 1999, and Semino 2002, among others). I exclude from consideration the influential narratological studies of Prince 1973 and 1982, Iser 1978, Genette 1980, Bonheim 1982, Fludernik 1996, Bal 1997, Herman 1999, Ireland 2001, and Bortolussi and Dixon 2003, since none of these focuses on the strictly text-linguistic.

On the other hand, important contributions have been made to our understanding of textual prospection, and narrative coherence and continuity, by a number of text and literary linguists and corpus linguists (e.g., Sinclair 2004, Stubbs 2005, Hoey 1994 and 2001, Hardy 2005, and Emmott 1997; and in more general treatments, Carter 2004, Cook 1994, Simpson 2004, Stockwell 2002, and Semino 1997). This is a large topic with diverse implications, including those for the fundamentals of language theory, and many aspects of the narrative text's shaping of the reader's interpretive processing remain to be fully described and explained. What creates narrative interest is a profound issue with connections extending far beyond the verbal organisation of the text, into psychology, social history and culture; the present approach, rooted in linguistic and corpus analysis of the language of the text, addresses only one major contributory factor. The study falls broadly within what is coming to be known as 'corpus stylistics', which involves the rapid searching and sorting of electronic versions of texts, often assessed comparatively against appropriate reference corpora, i.e., large comput searchable gatherings of texts. Some versions of corpus analysis emphasise quantitative factors, and can be important because of the unprecedentedly large samples on which the analyst's findings are based (Biber 1988; Biber et al. 1999; Short and Semino 2003; Hoover 1999; Zyngier 2008); in other uses, there is more interest in the ways that corpus analysis can underpin changes in our theories of text and language (and perhaps style), with particular emphasis on the collocational

A defensible way of beginning the study of readers’ judgements of narrative progression is to elicit (directly or indirectly) comments and reactions from them concerning the text that they are in the process of reading, or have just finished reading (see, e.g., Miall and Kuiken 1994). I have used questionnaires, diagnostic tests, and other reader-response methods (both ‘online’ and ‘offline’, oral and written), and have sometimes used manipulated texts (e.g., with narrative sections re-ordered, or with changed or deleted wordings) to prompt readers to disclose their chief narrativity reactions. Such methods date back at least to reader-response studies of twenty-five and more years ago (Dillon 1978; Kintgen 1983), not to mention the earlier work of I. A. Richards (1929), updated in various ways in psycholinguistic studies of reader processing, and empirical studies of literature (Ger-rig 1993; Hakemulder 2000; Kreuz and McNeally 1996; Martindale 1990; van Peer 1986; van Peer and Chatman 2001; van Peer, Zyngier and Hakemulder 2007; Miall 2006, Miall and Kuiken 1998, 1999; Lohafer 2003, reviewed in Toolan 2005; and Zwaan 1993).

The corpus analytic methods used to identify and trace lexical and phrasal networks include resources available in Scott’s WordSmith Tools software (Scott 2004) and Rayson’s Wmatrix text-analysing software (Rayson 2007), which enable the user to draw up frequency or alphabetised concordances for texts, collocations, keyword lists, and frequent phrase clusters. I will also make some use of computer programs that help track distributions of recurring words and new words across a text, such as Youmans’s Vocabulary Management Profiles (Youmans 1994). However, these programs sometimes have to be supplemented by less automated searches, e.g. of all semantic associates and paraphrases of identified keywords, using online thesauruses and a collocations dictionary to assist in delimiting such lexical networks. In these areas of methodology, where attention to forms is supplemented by consideration of lexical semantics, a degree of subjective judgement is inescapable. Besides keywords, collocates and clusters, corpus methods can be used to highlight the occurrence and distribution of the lexico-grammar that is crucial to evaluation, such as modality, negation, irrealis statements (e.g. statements about imagined or hypothetical events or entities), and cohesive naming chains. Focussing on such a restricted array of linguistic phenomena in
the course of analysing narrative progression is deliberately selective: the objective is to determine whether such factors are critical to effects of narrative progression in the text and of expectation in the reader. But even a book-length study such as this can give no more than a part of the full picture, and it attends almost entirely to what is textually explicit, even though it is incontrovertible that what is textually implicit must also be taken into consideration. By the textually implicit I refer to what a reader derives from a text when they bring to it reasonable and relevant inferences, schema and genre expectations, and background or cultural knowledge. The view taken here is that the implicit, the ‘background’, can only begin to be described once the explicit has been attended to; the explicit text is our royal and only secure road to the implicit. Thus while a full account of in medias res narrative expectation needs to address the complex integration of implicit guidance with the narrative’s explicit lexical and phrasal patterns of information, a clearer picture of the latter is a precondition to a better understanding of the former.

In the following chapters I will introduce particular aspects of corpus- and text-linguistic methodology when first reporting their use. But at this point a fuller description of key ideas that are central to the book’s aims may be useful. These ideas concern the text’s guiding of expectation, the ‘integrity’ of a text, and (in Chapter 2) collocational stylistics. In the course of discussing them I hope to situate the present study in relation to some of the most relevant previous work in these areas.

3. On genre, relevance, scripts, and background

The words on the page that we read and make sense of, find the story in, and derive expectations and feelings and understanding from, always come to us integrated within a context of situation and of culture. A variety of kinds of background knowledge informs our reading of a present narrative text, and prominent among these is knowledge of genre. The genres are like a mapping of the terrain of discourse into a number of mountain ranges; like the mountain ranges their exact extent is indeterminate, and like mountain ranges they can rise and fall over time (but quicker). Just as a lot of terrain is not clearly within one range or another but somewhere in between, a lot of texts are not unambiguously of one genre or another.

“it is because genres exist as an institution,” writes Todorov, “that they function as ‘horizons of expectation’ for readers and as ‘models of writing’ for authors” (1990: 18). Genres channel the reader’s inferences, help create intelligibility and coherence, and delimit the scope of interpretation. They have been explored from many perspectives in recent years (see, e.g., Duff 2000, Neale 1980, Fowler 1982, Cohen 1989, Gerhart 1992; of more direct relevance to linguistic studies are, e.g., Biber 1988, Swales 1990, Bhatia 2004, Martin 1992, Martin and Christie 1997, and
Threadgold 1989). Genres are normative, established, institutionalised types of discourse, often where a somewhat specialised purpose has emerged. They contrast with, for example, the relatively undirected and 'free' form of casual conversation, and with spontaneous or reactive interactional discourses generally. A genre is often described as a staged and phased distinctive organisation of discourse which accompanies a staged, goal-oriented social process; and different genres tend to be realised through somewhat distinct linguistic registers. In literature there are the macro-genres of the novel, the play, and the poem, and then more formally and culturally specific genres such as the crime novel, the revenge tragedy, and the villanelle; in sociolinguistics and discourse analysis, standardly-recognised genres include the personal narrative, the service encounter, the formal classroom lesson, and the doctor-patient medical consultation. In assuming that the discourse we are engaged in or the text we are reading is an instance of a particular genre (and no other), we are hazard ing a generally helpful pre-judging, which, strictly speaking, is a real-time or 'online' judging about the broad likely purposes, progression and outcome of the present interaction. So genre knowledge certainly shapes some reader expectations when a person begins reading a narrative, and needs to be acknowledged here. At the same time it is fair to say that, confronted with something that appears to be a 20th century short story, even the well-read reader's genre-based expectations will be quite general: the reader will expect to be told or shown how something will happen to someone, somewhere, but quite who and what and when and where are all open possibilities, to be delimited within the text alone. Anything can happen. Still, the genre-based expectation is that there will indeed be a particular thing that happens, to one or more individualised characters, in a particular spatio-temporal situation, or several.

In other ways, too, the reading and making sense of any narrative is guided in part by material other than the text itself. One part of this guiding background is often called world or encyclopaedic knowledge: the reader's knowledge of whichever of the innumerable specialised fields – the sciences, the arts, the crafts, religion, history, languages, botany, cuisine, medicine, law, graphic design, and so on – is relevant to the particular text being processed. Another part of the background is not public and general knowledge, but the reader's personal values, belief, or ideology. These may largely be shaped by supra-personal forces such as nationality, ethnicity, gender, class, and age, but both the personal values and the social forces will interact with our more personal experiences so that unique personal response, on the basis of the individual's personal experience and values, is inevitably involved in text processing. Just as no two people need respond to a very short public text in the same way (consider a 'No Smoking' sign in a public building read by the anti-smoking campaigner with an inward cheer, but by the tobacco-user with a snort of contempt for this petty oppression), people inevitably see longer and
more complex texts through different values-coloured spectacles. A useful informal term for these personal values, less lofty than ‘ideology’ or ‘beliefs’ but fitting perhaps to our materialist age, is ‘baggage’.

The metaphor of baggage, such a mot-clef of the 1980’s and 1990’s, could be a useful way of thinking about the background of attendant conditions, processes, and entities (concrete or abstract) that we can readily imagine likely to be potentially manifest in a narrative episode, mention alone of which would not usually constitute foregrounding or noticeability. We could speak of the ‘normal amount of baggage’ with which the depicted person might make their daily commute to work, or take a one-week holiday, or permanently emigrate. This might begin with physical accompaniments, but can be extended to include emotional and social debts and obligations, familial expectations, and societal pressures.

Every literary narrative episode is susceptible to a ‘colouring in’ by the reader, entitled to embark on the forming of various implicatures of weaker or remoter relevance. The term implicature was introduced by the philosopher of language Paul Grice to refer to those inferences that a reader or listener can reasonably deduce – and may often be intended to deduce – from a text, when its overt content is assessed against a contextually-reasonable level of cooperativeness (see especially Grice 1975) This colouring or filling in can be thought of as a visualisation (of the scene, of the characters, of the point of it all), performed with the ‘aid’ of our irremovable values-coloured spectacles. But even visualisation here is a figure for something richer, a conceptualisation. There stands Eveline, the eponymous heroine of Joyce’s “Eveline”, at the quayside, frozen with terror at the huge step Frank enjoins her to take, and as readers we can visualise or conceptualise this in all the ways not explicitly excluded by the text. Beyond the textual guidance, only general standards of plausibility (often called naturalisation or vraisemblance: see Culler 1976), which come into play as soon as we voice our weird ideas to others, act as constraints on this mental visualising, coloured by our own values. So I do not see Eveline standing there tanned, in a bikini and high heels, like a girl out of Updike’s “A & P” story; nor is she in a wheelchair; nor is she sporting a nose-stud.

As for literal baggage, does she have any at all? As critics have noted, there is no mention of Eveline packing any bag during her long reverie, so is she at the quayside sans passport, toothbrush and nightie? We shall never know. The soldiers, by contrast, are all set with their brown baggages, the text reports. The example has become flippan, but it is mentioned as representative of the processes of visualisation and conceptualisation readers perform on texts all the time, drawing on their own values, genre-expectations, intertextual awareness and background knowledge. Mostly we are unaware we are visualising characters and their situations at all, until we are pulled up short by another reader’s incongruous remark (about Eveline having a suntan, or Lady Macbeth having many children). But all our
reading of narratives (fictional or otherwise) involves a sense-making enrichment of explicit textual evidence by details and connections that we ‘can’t help’ making or that we judge relevant or likely, drawing on our familiarity with seemingly similar narrativised experiences; in cognitive studies these have been called, tellingly, ‘scripts.’ The term is telling because even in attempting to characterise extra-linguistic mental activity, a label invoking written text has been used (see Schank and Abelson 1977, Goddard and Wierzbicka 2004). Thus most readers will have in mind a vague and general ‘script’ about some of the stages and material involved if, like Eveline, a young person permanently leaves home. These mental scripts are another context within which a reader interprets a narrative about someone leaving home, or holding a dinner party, or waiting in hospital for their child to recover consciousness after an accident.

4. Guided expectation

While acknowledging the importance to narrative reception of genre, background knowledge and values, and mental scripts, this book’s goal is to draw on text linguistic theory and corpus analytic methods to enrich the description of narrativity and of such reader reactions as suspense, surprise, secrecy, ellipsis, closure, and mystery. The issue is a fundamental one in the creation and understanding of text structure. What is going to happen next – and later, and finally – are the sorts of questions that readers continually ask themselves in the course of reading any text. The read text prompts us to develop expectations, although these are rarely formulated in words. Thereafter, those expectations inform a multitude of reactions and experiences. To mention two of these, delayed fulfilment of expectations causes suspense, while the abrupt overturning of them can lead to surprise. Questions about what is going to happen next and later are only more sharply posed in relation to narrative text than in relation to texts in general. And a fundamental premise of the present study is that the ‘what next?’ questions posed by the reader are given qualified answers from as early as the opening of a written narrative.

As soon as the reading of a story is under way, the reader begins to make a ‘mental model’ about many things in the unfolding text (what it is about, who the main characters are, what the main themes may be, what the setting is, and so on; cognitive psychologists call this the ‘situation model’). All those judgements must be made on the basis of textually-declared information, given on the surface of the text as it were, or on the basis of information inferable in the light of what has been textually declared. Otherwise, the modelling and interpretation are not sourced in the text at all, and are no better than a misreading. In “Bliss”, for example,
the story’s first few pages tell you that there is a woman named Bertha Young, that she is thirty years old, that she has a husband Harry and a baby, that the story-time is roughly Edwardian (c. 1910), that the setting is affluent, fashionable, mildly Bohemian London, among other pieces of information. But the reader quite naturally and inevitably goes beyond this kind of summarised sense-making, a synthesis of the text already read and the information derivable from it. The reader in addition makes predictions, which are informed guesses, about what will happen or be shown to have happened in succeeding pages. These reports about the narrative present and the narrative past amount to the same thing, from the point of view of the experiencing reader. Some readers will develop far more elaborate (and accurate or reliable) expectations than others, and some will be more able to talk about their expectations than others; but all proficient readers predict to some extent.

That expectation is a fundamental human propensity, invaluable to our ability to adapt to shifting circumstances, is generally recognised. But how we cope with ‘the expected’ relates directly to how we cope with the unexpected also, and as Chafe (1990) has remarked, no amount of reliance on schemas and pre-conceived modelling is truly adequate to all the changes and differences that we encounter with every new day. A similar conviction, connected to a view of human communication as continually subject to semiotic renewal, is a central theme in integrational linguistic theory (see Harris 1998). Chafe also notes that we even seem to welcome the unexpected, and the arousal of emotion that it triggers; the arousal, in turn, helps us to cope with that unexpected. As he thought-provokingly comments:

An unexpected input is a threat to our accustomed modelling of the world, and our primary reaction is to reject it by getting rid of its source. Excitement, anger, and aggression are essential to this rejection process. From this reaction comes the satisfaction people find in xenophobia, and the everyday satisfaction people feel in complaining about things that are unfamiliar…. Ironically, we are not content with inputs that conform to our expectations, but prefer experiences that contain the unexpected, that arouse us to reject their source. Rejecting the unfamiliar is what we do best. From this follows the fact that we are satisfied only with narratives whose point is some kind of conflict with expectations…. Narratives that present a conflict with expectations provide excitement by exercising the mind’s innate capacity to react and deal with inputs of such a kind.

(Chafe 1990: 83)

Of course Chafe is not suggesting that the only reason we visit narratives is for the pleasure of rejecting their ‘unexpected inputs’; besides, the fictional narratives of short stories mean that any ‘rejection’ readers enact will be vicarious or notional. What Chafe’s remarks particularly suggest is that the reader’s processing of narrative is emotional as well as intellectual, and attuned to recognition of the expected and the unexpected, without which rejection and perhaps other reactions cannot emerge.
5. Text integrity

Text-predicting is a fundamental assumption of this study, and can be contrasted with an account of text and text-processing in which the possibility of ‘zero prospection’ is allowed. My assumption is that a text with ‘zero prospection’ – that is, one in which at any non-final point in the text a reader has no idea whatsoever as to how the text will likely continue and finish – is zero-coherence discourse and therefore not text at all, but merely a sample of randomly-selected and randomly-combined sentences. To read two sentences in sequence but without any sense or expectation of the second being in some sense prepared for and made relevant (directly or indirectly) by the first, is to read those two sentences as not-a-text, i.e., as two separate texts (or fragments of two separate texts).

We do this all the time in everyday life: the commuter looks out of the train window at an advertising billboard telling her that BA flights to Brazil are now cheaper than ever, and then returns to reading her newspaper story about some government minister in trouble: “Clarke will make a full statement in the House of Commons on Tuesday”. If the commuter’s reading is now interrupted by someone wheeling a refreshments trolley who leans towards her and says “Any teas or coffees at all?”, then this, too, will be rightly interpreted as unrelated to the BA or the Clarke sentences. These three sentences, two read in sequence, the third heard, do not comprise a single text, in the sense of interest here. They are fragments or wholes of three distinct texts, three distinct communicational events. Our ability to distinguish separate texts is invaluable. In the scenario described, for example, it helps the commuter understand that the offer of tea or coffee is not some kind of summary of the full statement Clarke will make on Tuesday. This is not to deny that the three sentences can be made part of a single text, given a radical change of footing and a particular communicational framework. For example, in a personal narrative of a journey to work: and then I saw a billboard saying that…. and then I read that Clarke will make a statement… and then the refreshments person came along and said “Any teas…?”.. But it is hard, even within the personal narrative framework, to deny the largely arbitrary juxtaposition here of different ‘speakers’, different interests, a randomness of the progression, and lack of a perceptible multi-sentence design.

This rough and ready segmentation of the communicational and interactional flow of experience is an invaluable resource for language-users. It confirms the important fact that we take some sequences of sentences to be intrinsically-related text, while we class other sequences as fairly reliably not. And certain uses of language (such as the desultory talk, too discontinuous to count as conversation) that people produce in the course of the working day while engaged in other work and other kinds of discourse may be borderline between text and non-text.
6. Predictive reading

One means of making more palpable our quite disciplined powers of expectation, driven by an integration of textual signalling and background knowledge, is to present readers with all but the final line of a relatively short poem which has a narrative element. Consider, for example, the following poem by Philip Larkin, from which the final line has been removed:

‘Take One Home for the Kiddies’
On shallow straw, in shadeless glass,
Huddled by empty bowls, they sleep:
No dark, no dam, no earth, no grass –
Mam, get us one of them to keep.
Living toys are something novel,
But it soon wears off somehow.
Fetch the shoebox, fetch the shovel –

I have asked readers what they calculate the final line might be, and what means they use to arrive at a suggested ‘solution’. In fact, can the reader in large degree predict the final line, almost to the point that they might, uncannily, feel they have already encountered it, before reading Larkin’s actual version? It emerges that readers rapidly make a number of guiding assumptions. The final line, they predict, is quite likely to:

– begin with *Mam*
– use a first person plural pronoun (*we/us*), just as the final line in stanza one does
– report something that either Mam or we kiddies is/are doing (present tense)
– end with a word that rhymes with the *how* of *somehow*, which may be temporal to match the *soon* of the partner line, and is very likely to be monosyllabic, being the final word in the final line
– this doing is poem- and story-final, and done to something treated as a living toy where the novelty has worn off, involving a shoebox and a shovel: background knowledge suggests a burial, which requires the inference that the ‘living toy’ is now dead, and this is consonant with the wretched conditions described in the previous seven lines
Armed with the above guidance and expectations, readers frequently suggest a wording of the final line such as *Mam, can you bury them now?* or *Mam, we're having a funeral now.* Neither is far removed from Larkin's actual construction:

Mam, we're playing funerals now.

The enforced comparison between candidate endings and Larkin's actual one is itself instructive. Readers see that *playing funerals* is not only rhythmically but also in terms of word choice more fitting than *having a funeral* or *giving it a funeral.* Larkin's *playing* associates – collocates – with *toys* in line 5, and probably with *kiddies* in the title. A theme of trivialising, of using a pet rather than caring for it, is extended to its sobering end. This amounts to saying that, interacting with the schemata and scripts or other background knowledge that those lines seem to draw upon, *we're playing funerals* meets the expectations created by the foregoing lines better than *we're having a funeral* would. *We're playing funerals* more fully meets the prospection created by the plausibly-interpreted previous lines.

7. Conclusion

I have used the subtitle ‘narrative prospecting’ for this introductory chapter because that does sum up both what this book aims to do and what I believe readers processing a literary narrative do. The corpus analyst is something like a prospector, too, in the sense of one who sifts through rocks and dirt, looking for a seam, for nuggets, a pattern of collocating wordings, an association of a texture or colour or malleability with a particular valued material: *oil! gold!* or, in the present case, *expectation-generating narrativity!* There is some similarity with the automated sifting and sorting of text undertaken in many commercial fields, known as ‘data mining.’ Interestingly, Moon has used a similar figure in a recent paper, describing her work on the stylistically distinctive vocabulary of particular texts and authors as a boring down through the strata to rich seams of lexicosemantic significance (Moon 2007).

But readers, too, are prospectors, in the way that they *take* the signalling, the material of the text, and use this to judge the narrative arc of what has been already told, and to form expectations of what is passing and to come. In attending to the foregrounding and patterning that are constitutive of the text, they are as purposeful and cognitively engaged as any prospector or engineer.

This book explores an issue of deep interest to readers: how it is that great writers fashion stories in ways that lead us on, sometimes lead us astray, draw us in, take up all our powers of attention and concentration, and induce in us the whole gamut of reactions and emotions – including desire, revulsion, inspiration, grief,
and fear. But the book does not aim to ‘demystify’ or ‘solve’ this complex achievement. Another premise of this study (I would claim it is an enabling, not a disabling one) is that no full and final explanation of narrative expectations is logically possible, just as there is no logical reason to suppose that the writing of short stories or other artistic production should ever come to an end, on the grounds that all forms have become ‘exhausted’.

The next chapter supplies necessary background on recent studies of prospection and lexical patterning which have influenced the present study, together with a brief discussion of some basic methodological issues.
1. Studies in discourse prospection and expectation

In this chapter I briefly introduce some of the most important scholarly influences on this study. Several ideas only mentioned here are adapted and applied extensively in subsequent chapters.

The general conviction that earlier sentences in a narrative text shape the expectations we carry forward to our reading of later sentences has a long history. In the broadest way possible, the topic begins with the Aristotelian statement of first principles, that a story must have a beginning, a middle, and an end. It is our ability as writers and readers of text to know that we are in the middle of a story, that its beginning has already finished and been absorbed, and that its end is 'ahead'; that enables more specific expectations about the ending to be entertained. In the modern period, the study of English textuality was greatly advanced by the work of Halliday and Hasan (1976) and Hasan (1984, 1985) on what they call the cohesion in texts – the varied grammatical and lexical resources, including reference, ellipsis, conjunction, and lexical patterning – that help bind sentences together into an economically-composed unity (other important early contributions included Winter 1994 and Hoey 1983).

Another warrant comes in the idea of ‘adjacency pairs’ propounded by Conversation Analysts, who have discussed the strong ‘preference’, in conversational interaction, for requests to be followed in timely fashion by a complying act (or, as a second best, by some account of why the addressed party cannot or will not meet the request); for offers and invitations to be followed, again preferably, by an act (verbal or otherwise) of acceptance; and for questions to be answered by informings. The structured sequentiality of many kinds of discourse was influentially discussed by Sinclair and Coulthard (1975) in terms of ‘prospection’. This pioneering study showed how a teacher’s ‘instructional’ question in the traditional British classroom requires and expects an answer from one or more pupils (there may be an intervening ‘bid to answer’ and ‘nomination of answerer’ sequence) and, crucially, that a third segment is also expected in which the teacher evaluates or gives feedback on the answer proffered.

Indeed many strands of Sinclair’s work reflect his interest in the ‘partial predictability’ of the textual sequence. A corollary of that predictability is that,
generally speaking, there is less continual (and effortful or costly) choosing going on in our creation or interpreting of discourse. Within the clause or sentence, for example, Sinclair influentially argued that the composer is often not tasked with selecting from an open paradigm of equal-weight alternatives as a phrase unfolds (pick any article; now pick any adjective; now pick any compatible noun). Often, we speak and write in “semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments” (Sinclair, 1991: 110). By the same token, as readers and addressees, we can expect that what we hear or read will often be holistically-interpretable pre-fabricated multi-word phrases. Sinclair calls this chunking tendency ‘the idiom principle’ and contrasts it with the more familiar and dominant ‘open choice’ principle, the picture of sentence-construction as a continual choosing of items to fill slots, constrained only by considerations of grammaticality. It is important to recognise, however, that Sinclair is not rejecting the open choice principle: he sees both principles at work, in tension, in language production.

Textual progression has been an enduring interest of linguists at Birmingham University, committed to a form-oriented discourse analysis, as the following quotation from Sinclair (2004: 13) underscores: “The more that attention has been focussed on the prospective qualities of discourse the more accurate and powerful the description has become.” Complementary to Sinclair’s notion of prospection is his theory of encapsulation in text, which holds that any kind of discourse is designed in such a way that a first segment of text raises implicit questions that subsequent discourse will address, wholly or in part, and that each succeeding segment of text may open or prospect further kinds of question or incompleteness for yet later text to resolve (Sinclair 2004, especially 82–101 and 115–127).

A recent contribution which draws on aspects of the above approaches but goes much further in proposing a general theory of expected co-occurrence at every level from the lexical to the generic, is Hoey’s theory of lexicogrammatical priming (Hoey 2005). Adopting and elaborating on the psycholinguistic notion of priming, Hoey argues that for a significant number of quite frequent lexical items in a language (his examples are from English), fluent language users have strong tacit expectations, reflected in extensive discourse evidence, about the words with which those items tend to co-occur, the sentence elements and semantic roles those phrases or word-sequences tend to fill, and even the genres they tend to occur in.

Hoey’s study is an original and ambitious corpus-based elaboration of collocation theory, the latter being an approach to language in use in texts which has burgeoned since some seminal papers by Firth in the early 1950s (Firth 1957). At the heart of this approach is attention to the utterance-structuring that is rooted in the tendency of certain words to co-occur in sequences on grounds of mutual attraction or compatibility rather than purely structural permissibility. These
co-occurrence tendencies or primings foster predictability and ‘naturalness’, but to be primed is by no means to be compelled. The clustering or mutual attraction of words is not of the quasi-automatic kind arguably found in idioms and fixed expressions. Many of the latter are so fixed, invariant, and resistant to decomposition as to be almost classifiable as multi-segment single words. Collocational priming is a looser bonding of discoursal chunks such that, when we hear or read part of that chunk, the remainder is expectable or at least entirely unsurprising. And the chunks are also easy enough to interpret, provided one is otherwise familiar with the words involved, whereas many traditional idioms – *kick the bucket, blow the whistle on,* etc. – cannot easily be understood merely on the basis of familiarity with their constituent words (see Makkai 1972, Fillmore et al. 1988: 504–5).

Hoey’s is just one model of textual patterning based in collocation theory, but it is one of the most developed currently available (but see also the work of Biber, and the notion of ‘lexical bundles’: Biber et al. 1999, Biber et al. 2004). Hoey asserts that “We can only account for collocation if we assume that every word is mentally primed for collocational use” (Hoey 2005: 8). And he adds:

As a word is acquired through encounters with it in speech and writing, it becomes cumulatively loaded with the contexts and co-texts in which it is encountered, and our knowledge of it includes the fact that it co-occurs with certain other words in certain kinds of context.

(Hoey idem: ibidem)

Prior to the 2005 publication, one of Hoey’s major interests had been in patterning by repetition within and across the text, and in how the patternings (and variations in patterning) reflected such fundamental phases of a text as Setting, Problem, Solution attempt, and Evaluation (e.g., Hoey 1983 and 2001). The new focus is not such a radical departure from the emphasis on repetition as it might first appear. This is because priming is an entrenched (mental) awareness of word co-selection repetition, and one that seems sensitive to mode, variety or genre, rather than operating indiscriminately across all uses of a language. In fact, Hoey goes further than this to say that simple repetition is itself a form of collocation: it is a word collocating with itself.

A priming theory raises many issues that await fuller answers, for example concerning the degree to which it is stronger or weaker for different users of a language, and the consequences of large variations in its strength. If it is indeed stronger or weaker (in different contexts) for different speakers/readers, what might be the consequences (and causes) of having acquired only very weak primings? Are there even some reasonably fluent users of English (non-literate, for example) who acquire very little priming awareness? At the same time, why should we say that the individual word is primed for collocational use, for the words it
tends to keep company with, and not say that individual choices of word (and of spelling, and of pronunciation) are also a matter of priming? What is the distinction, then, between what some have called ‘linguistic knowledge’ or ability and what Hoey is now calling language priming?

The present study takes some inspiration from Sinclair’s ideas about the idiom principle underlying the mutually-selecting constrained choices in the multi-word lexical item, and from Hoey’s ideas about how earlier words, phrases, and textual segments prime the occurrence of later ones. There is some overlap in their approaches, but Sinclair’s work focuses on multi-word chunks and is strictly textual, while Hoey has a more ambitious theory about expectable co-selection at all levels of textual analysis, and claims that this is as much a cognitive as a textual phenomenon. An enquiry into ‘what is expected to follow’ in a short story (i.e. what is normal or usual and known by readers to be normal or usual) is then a huge magnification along the text’s syntagmatic axis of what grammar pattern studies embark upon. Potentially, the reach or range of some prospections is not just across a few words, but across very many sentences, to the end of the text.

2. Collocational stylistics

What is collocation? A simplest definition is that it is the lexical company a word tends to keep (a minority of which co-habitings give rise to clichés). Sinclair has characterised collocation as “the co-occurrence of words with no more than four intervening words” (2004: 141); from this we can postulate that two (or more) words are collocates when they co-occur, noticeably often, no more than five words apart in the same sentence. But for such an intuitively straightforward idea (we drink strong tea, not powerful tea, drive through dark nights during which animals on the road may be scarcely visible to the naked eye, and this may cause an accident – indeed, a tragic accident, etc. etc.) it has prompted a striking variety of different characterisations, as Hoey (2005: 3–5) shows in a brief review. His own definition is that collocation is “A psychological association between words (rather than lemmas) up to four words apart and [is] evidenced by their occurrence together in corpora more often than is explicable in terms of random distribution”. (5)

This is careful in many ways, and strives to avoid what Hoey sees as the imprecise claims of some previous formulations, but it is itself open to question in its assertion that collocation is essentially psychological, with more-than-chance textual co-occurrence as merely evidential. Without neglecting the psychological
dimension, our definition of collocation should not subordinate the textual phenomenon, and therefore I would propose the following:

Collocation is the far-greater-than-chance tendency of particular words to co-occur (adjacently or within a few words of one another). These co-occurrence tendencies have text-constructional and semantic implications. Proficient native language users are equipped, by their communicational experience, with the commoner collocational tendencies and implications.

Like Sinclair before him, Hoey associates collocation with ‘naturalness’: phrasings that a native speaker would regard as normal, fluent, and proficient. Non-natural phrasings, insensitive to or neglectful of collocation norms, tend to be judged non-fluent or clumsy: they are ‘correct,’ coherent and meaningful but not fully satisfactory. In some respects they look or sound rather non-native (phrasings of the kind found in ‘advanced foreign-learner language’). Hoey supports this distinction by taking a Bill Bryson sentence which is said to flow easily and naturally thanks to its numerous interlocking collocations:

*In winter Hammerfest is a thirty-hour ride by bus from Oslo, though why anyone would want to go there in winter is a question worth considering.*

Hoey shows us what happens if this is rewritten as a sentence that is equally grammatical and meaningful, but quite awkward:

*Through winter, rides between Oslo and Hammerfest use thirty hours up in a bus, though why travellers would select to ride there then might be pondered.*

The conversion has removed all the smooth collocational phrasings, the primings that make Bryson’s sentence an easy, natural read.

Collocation is often mentioned in the same breath as colligation, but the two are distinguishable. Colligation is the tendency not of a few particular words, but of any items from an entire grammatical sub-class, to co-occur with a specific lexical item. Butler (2004: 154) comments: “Originally coined by Firth to mean the co-occurrence relationship between grammatical categories such as noun and adjective, [colligation is] now used more widely to cover relationships between grammatical categories and particular lexical words as well.” While Firth was one of the earliest to begin to theorise these lexical co-occurrence tendencies, several other commentators on the language were aware of the phenomenon in their own way, including Hornby (1954), and Fowler (1965), and such writers as George Orwell (1968 [1946]) and Flann O’Brien (1999 [1968], writing under his journalist sobriquet of Myles na gCopaleen).
According to some corpus linguists, Firth’s own exposition of collocation has proved the most enduring and inspiring of those that have been proposed, and the most clearly confirmed by computational testing. In his words,

Meaning by collocation is an abstraction at the syntagmatic level and is not directly concerned with the conceptual or idea approach to the meaning of words. One of the meanings of night is its collocability with dark…

(Firth 1957: 196)

And yet Firth is not always the most explicit of guides. What exactly is meant by ‘the conceptual or idea approach’ to word meaning, said to be distinct from meaning by collocation? The precise nature of the distinction Firth had in mind remains uncertain for many commentators, who have also noted how Firth gradually adjusted his view of ‘pre-fabricated’ lexical phrasing (see, e.g., Joseph 2003). Firth’s point in saying that collocation is ‘an abstraction at the syntagmatic level’ may be that while, for instance, dark and night are frequently collocates, in that order, this is a tendency and not a rule or obligation. By contrast the ‘rule’ in English noun phrases that adjectives precede nouns, and determiners precede both these, might amount to ‘concrete syntagmatics’. Collocational tendencies are removed from such general rule-like patterns, which apply to very large sets of words (‘all animate-referent nouns’ for example) with limited regard to their particular meanings. Collocational patterns as between dark and night apply for much smaller lexical-cum-semantic sets – single word pairs being the limiting case but more often, as Stubbs (2001) has emphasised, clusters of semantically-related items in each collocate position.

How can (or might) collocation studies bear on stylistic analyses of narrative progression? Perhaps no linguist has been more enthusiastic about the potential of collocational and corpus linguistics to reform stylistics than Bill Louw. In his recent papers Louw has been critical of current stylistics, but argues that collocation and corpus linguistics can effect a rescue:

If there is any branch of linguistics likely to bring about the emancipation of stylistics, collocation will be that discipline. Once the contribution of collocation to the act of reading has been fully documented we may well discover that collocation becomes the key to meaning. In corpus and computer-based terms: collocation has virtually become instrumentation for language. (Louw 2006: 92–3)

Overstated though this may be, Louw has given a glimpse of collocational instrumentation with his influential but still-controversial idea of ‘semantic prosody’. Louw uses this phrase to refer to the tendency, in certain node-collocate phrases, for a semantic association (usually a negative evaluation) of a common collocate to transfer to the pragmatic meaning of the node and other collocates as used in all
manner of contexts. Just as a phonological prosody can ‘wash’ over a whole phrase or line, a semantic prosody is not confined to the segment that is one word, its notional source, but colours or infuses adjacent words. And this semantic colouring or prosody is of a kind that does not self-evidently inhere in the intensional meaning of the core or node item, but is a “consistent aura of meaning with which a form is imbued by its collocates” (Louw 1993: 157). The primary function of a semantic prosody is “the expression of the attitude of its speaker or writer towards some pragmatic situation” (Louw 2000: 57). The verbal phrase set in, so the argument runs, does not in itself necessarily mean something ‘bad’ took hold. But we so often report that the bad weather set in that you cannot easily use set in sincerely with a positive thing. In fact, Louw proposes that a fluent speaker using a negative semantic prosody word (utterly, set in, cause [as a verb]) in a positive phrase should be suspected of either irony or insincerity. To give an (invented) example of the exploitation of semantic prosody for ironic effect, if I find Great Aunt Jemima’s heartiness overpowering and report that “The family holiday was very quiet at first but then Great Aunt Jemima arrived and round-the-clock jollity set in”, the alert addressee will calculate that I found the jollity unwelcome.

Louw believes that the bottom-up objectivity of collocational studies can enable it to uncover deeper insights into the phrasings in literary texts than those of the best intuitive critical reading. This may be a longer-term ambition, but in the meantime a more realistic but still difficult goal is the uncovering of some of the textual patternings that give rise to rich critical readings. While pursuing that goal it may be unwise to characterise the analytical procedures in ways that endow the phenomena or methods with too many of the powers that are actually wielded by individuals and groups. It takes a human analyst to uncover insights and disclose things, using collocational analyses; it is humans who actually do these analyses, not the relatively simple but rapid matching and counting routines that recent computing resources hugely facilitate. Nor is the claim that corpus linguistics is a methodology universally recognised. Teubert, for example, argues that “Corpus linguistics is not in itself a method: many different methods are used in processing and analysing corpus data. It is rather an insistence on working only with real language data taken from the discourse in a principled way and compiled into a corpus.” (Teubert 2005: 4). Equally, there are dangers of misrepresentation of methodology if we claim, without qualification, that computer programmes can count words and display textual examples and reveal statistically significant patterns of distribution. The software counts words the way a garden fork digs an allotment: we should beware of conflating the instrument and the agent.

Similarly computer tools do not strictly organise and display textual examples – certainly not remotely analogously to the way a person can organise and display examples of anything. If we remove the nominalisations and anthropomorphism
with which we endow them, computer hardware and software are simply tremendously helpful resources with which human analysts can do certain kinds of counting, comparing, displaying, etc., of certain kinds of material. By comparison with the old methods in which the solitary scholar read and listened, transcribed and tape-recorded, drew up lists and counted examples in fives, the computer has brought huge changes and massive advances, with the gains far exceeding the losses. What is immensely more feasible today, for anyone armed with a computer, is a degree of manipulation of the chunks that make up texts, their rearrangement and comparison, fostering particular kinds of attention to particular kinds of pattern hitherto difficult to perceive. But by the same token there is, as inevitable corollary, the risk of neglecting other kinds of pattern that were previously attended to more fully. Mahlberg rightly notes that “A corpus linguistic, bottom-up description of language prioritises lexis” (2006: 371); but it is reasonable to suppose that it does so since computers can so easily and reliably count and match innumerable space-bounded digitised forms (but can only ‘see’ and count something like syntactic embedding with much greater difficulty). Computers have enabled us to see texts, and language, differently. By comparison with the old so-called ‘manual’ computational methods for grappling with large amounts of linguistic material, it is perhaps analogous to displacement of the handfork by the diesel-powered tractor in the cultivation of 100-acre fields. But a tractor is not a farmer and corpus linguistic programmes do not do analysis.

In his contributions to corpus stylistics, Louw argues that corpus methods must not merely confirm intuitive literary critical and stylistic practice but overturn or displace it. There must be a “Malinowskian revaluation”, he argues (Louw 2006, 2007), in which linguistic analysis including stylistic analysis must finally become scientific, reading particular literary texts against a large sample of the whole language from which can be derived “general laws which have predictive value”. By way of qualifying this, and the particular notion of replicable objective science that it assumes, we may question the appropriateness of so extreme a contrasting of reliable objective collocational ‘meaning’ and unreliable subjective reader meanings. The asserted opposition between corpus evidence and reader intuition is in danger of overstatement, to the detriment of the larger theory of language. Rather, the value of corpus linguistics is and should continue to be that it brings into the spotlight patterns and tendencies of which the language user may have been unaware, or was in no position reliably to confirm, but which upon reflection they see to be the case, probable, and a telling norm.

In short, I do not see corpus linguistic findings as quite so revolutionary, and human language users as quite so misinformed by their intuitions, as some have claimed. On the contrary, it is rather important for my ways of proceeding that there not be a profound gulf between what our analytical methods highlight as
significant patterning and what readers by more informal and holistic means take from texts. The narrativity patterns explored in this book must not be so profound or hidden as to be beyond the attention of the reasonably careful reader; otherwise, claims about their subtle instrumentality in guiding the reader can be too easily dismissed as psychologically implausible.

Without doubt, one of the clashes between corpus linguistics and literary linguistics stems from the fact that corpus linguistics has been predominantly interested in the typical, and is often said to be at its best when it focuses on repeated occurrences and ignores the occasional exception, while literary linguistics is often interested in single texts each of which is to a degree unique and whose value inheres in their unique combination of elements. The uniqueness of each literary text is normally underwritten by its having a title, being published, sometimes in several editions, having an identified author, by considerations of copyright and intellectual property, and so on. Relatedly, I contend, the individual literary text (Great Expectations, “The Love of a Good Woman”, Waiting for Godot) is not a sample or ‘typical’ in the way that one business meeting, or one science journal abstract, from the discourse analytic perspective, is defensibly regarded as a sample from a larger collective practice or genre. This remains true even where the particular story is part of a book-length story collection, such as Joyce’s Dubliners, or Anderson’s Winesburg, Ohio. We do not, in the act of reading, treat any of the stories within those collections as samples of the whole, even if they are parts of the whole (each story is a complete part, not a typical part). The relation of each story to the whole collection is meronymic (x is part of y) rather than hyponymic (x is a kind of y).

This view clearly raises some challenges to the standard notion of genre, since it amounts to saying that the individual (classic or enduring) literary text is not ‘governed’ or rendered predictable by its genre-affiliation (even though there is no denying that literary stories can be usefully viewed as assignable to genres and sub-genres of short story). The contrast should not be overstated. There are some respects in which the relations between, for example, newspaper editorials and the individual editorial on the one hand, and modernist short stories and the individual story (Mansfield’s “Bliss”, for example), are quite comparable. But there are at least differences of degree, also. The single editorial tends to be more reliably a true sample, representative of and informative about editorials in general; but with regard to the detail of texture and lexicalised effects of interest here, “Bliss” is not a sample of the modernist story, even though it is an instance. Readers of the literary text tend to assert comparatively greater uniqueness of it than is expected of the non-literary one, and as a result less reliable typicality.

At this point what appears at first to be a minor terminological issue can be clarified. In the textual analyses in subsequent chapters, I will follow the lead of most collocation theorists in referring to chunks or segments of wording (“extended
lexical units” is one of Stubbs’s suggestions: Stubbs 2001) rather than simply to phrases. This is because the term ‘phrase’ is strongly associated with traditional grammar and syntax, and these in turn are predicated on a model of sentence-construction as a structured dependency of required and optional elements (e.g., a Verb Phrase and a Subject Noun Phrase, respectively), with a governing nuclear item as the ‘head’ of each phrase that is permitted and generated. In that traditional picture there are items (from lexical or phrasal categories) and rules governing their structured combination, but a wide-ranging openness or freedom is assumed as to possible combinations, except where some very general semantic restrictions apply. Any preposition can be followed by any noun phrase, and any transitive verb can take any noun as Object. Where exceptions need stating, they strive to be general (part of grammar, not of individualised lexis). Thus to the rule just stated about verbs and Object nouns, a generalised exception disallows – ordinarily and nonfiguratively – any verb entailing a human agent and patient from taking a nonhuman Subject or Object. The generalised exception is stated as a semantic rule (which must interact with the syntactic rules), and invented sentences breaching the rules are marked with an asterisk to advertise their ungrammaticality: *The dog contradicted the tree; *The car persuaded the parking meter. Collocation linguists see things very differently, however, since they focus on the enormous amount and variety of co-occurrence patterning of words. That evidence suggests that language production is often and normally not a matter of word-by-word or phrase-by-phrase rule-governed open choice. Instead, Sinclair’s idiom principle (noted earlier) is regarded by many collocation linguists as “at least as important as grammar in the explanation of how meaning arises in text” (Sinclair 1991: 112). In these ways, a seemingly minor terminological adjustment – from ‘phrase’ to ‘chunk’ – reflects a deep paradigm shift.

A final point in this section is one that should come to the fore again in Chapter 9, in which I attempt to test the semi-automatic large-scale detection of quite specific textual features that will be singled out for attention on linguistic and narratological grounds in Chapters 7 and 8. In my view, corpus analytic methods are only the tail, while one’s theory of language (or, here, of narrative text) remains the dog. Even where collocation is elevated as a shaping phenomenon, this must be finally on theoretical and not merely methodological grounds: we are not interested in collocation merely because it is now so much easier to examine. Like many other analysts now using corpus linguistic methods, I am aware of and resistant to the possibility that corpus linguistic tools could be a straitjacket. Stubbs notes there are at least three stages to a corpus stylistic analysis: “The linguist selects which features to study, the corpus linguist is restricted to features which the software can find, and these features still require a literary interpretation” (Stubbs 2005: 6). Commenting on this, Moon emphasises the potential problems caused
by the affordances and restrictions of the software: “By only using one type or form of software, it is only possible to retrieve certain categories of information…. Are there other categories which might be meaningful and useful, and other ways of presenting or foregrounding information which might lead to further interpretations?” (Moon 2007: 3). The only way to begin to answer this open question is by devoting considerable attention to the informational categories with which the text (here, the short narrative) can be analysed, that is, by critical review of the linguistic description and theory that have been advanced as explanatory of that type of text. In these ways this study will be based in corpus linguistic methods but in fact driven by textlinguistic and narratological theory.

3. Choosing an appropriate comparator corpus

Several of the previous sections have discussed various backgrounds (personal values, genre-expectations, scripts and schemata, real-world knowledge) that inform a reader’s processing of a narrative text. But this book particularly focusses on the putatively narrative-advancing lexical patternings in stories, using a reference corpus as the lexicogrammatical ‘background’ against which a story’s particular deployments of lexis can be treated as marked and foregrounded. The reference or comparator corpus, then, is used as, and therefore needs to be, a large sample of relevant language in normal, varied, and representative circumstances.

What texts are suitable candidates for such a reference corpus?

A first move in addressing the question is to decide to include in the reference corpus only texts that are themselves literary fiction, with a preponderance of short fiction. This reflects the now widespread recognition that to get at the detailed disposition of extended lexical units, corpus studies need to be genre-specific: an item’s patterning or prosody may be exclusive to a particular genre, and absent from that item’s behaviour in other genres. Somewhat similarly, corpus linguists have found that KWIC (key word in context) studies need to be word-specific and not merely lemma-specific: apple may well collocate, colligate, associate, and have a discourse prosody rather distinct from that of apples.

But there are considerable difficulties in attempting to specify what Austen invoked and Fish deconstructed as the ‘normal circumstances’ in which words are used. For the purposes of making general descriptive statements that might ultimately be relevant to language teaching, Tognini-Bonelli has plausibly argued that the data in a corpus must be assumed to be “taken from genuine communication of people going about their normal business” (2001: 55), and gives examples of data that in her view should be excluded from such a corpus as they reflect “some special restriction”: Bible translations or lyric poetry or interviews with psychiatric
patients (55). But even after such a screening process, the data may be unreliable since “writers and speakers sometimes deliberately break the normal conventions of an encounter… to illustrate a mistake…or for a number of other everyday reasons” (56). It is the tension here, between wishing to confine the data used to the normal and everyday, and then the acknowledgement that discourse can be norm-breaking in everyday usage for everyday reasons, that is of interest. Part of what drives corpus linguists’ anxieties over corpus design is the fear that any findings will ‘lose generality’, where they arise against the backcloth of a skewed or unrepresentative corpus. A similar anxiety arises in corpus stylistics applied to literary texts too, despite the fact that finally it is a desideratum that the description should lose generality, and should discriminate what is distinctive and unique in each of the texts under scrutiny. It seems that for every distinct study, a contextualised determination of what kind of data would count as genuine communication in the course of normal business needs to be made.

Choice of reference or comparator corpus remains a contentious matter, and differences in the selection of the larger corpus can have serious consequences for what emerges as key, or is overlooked as non-key, in the comparison. How ‘psychologically natural’ should the corpus be? Should the comparator corpus for a Raymond Carver story be all the other fiction by Carver, on the grounds that one of the more likely predictions we can make about someone reading “Boxes” is that – whatever else they may or may not have read – they are likely to be familiar with some other Carver stories? Or should the reference corpus be derived from fiction (in English) written (as best we can tell) in the year (or years immediately prior to) when “Boxes” was written? Should it be American fiction only, and/or short fiction only (in case, as some surmise, short fiction is lexicogrammatically distinct from the novel)? In this area there are dangers of distraction by excessive fastidiousness, especially since it seems that each choice carries its own weakness. Thus one might want, when analysing Carver stories, to exclude from the corpus any fiction written after Carver’s death, on the grounds that contemporary discourse with its blogging and podcasting and even, shocking to think, e-mailing, was no part of Carver’s linguistic world and could make for a false comparison. But the contrary view would note that just such new words and concepts are part of the background for most present-day and future readers of “Boxes”. Such cultural and technological changes, from a (recent) past discourse to a contemporary one, should not be reflected in the positive keywords of that previous-era text, except where one device has been entirely abandoned (e.g., telegrams) or directly supplanted by an alternative (as, perhaps, in the case of L.P’s displaced by CD’s).

The precise nature of the reference corpus requires particular care, for example, if the analysis is directed to a story’s top keywords, i.e., those words in the text that are disproportionately frequent (see Chapter 3). Imagine a story set in southern
England involving a character called Willetts; in the circumstances, the forms London and Willetts might both occur frequently in the target text, although the reader might rapidly sense that the mentions of London were only a maintenance of setting, while the name Willetts was central to the story’s narrativity. Provided that the reference corpus has some instances of London (and few or none of Willetts) then, to the analyst’s satisfaction, the latter term and not the former will be likely to figure in a listing of the story’s top keywords. But if the reference corpus of fiction happens to contain no mentions of London, then London would emerge as no less key than Willetts.

Much of this boils down to making some careful judgements as to what the background of a particular story or text can be reasonably said to be. Is the background against which a Carver story is most ‘naturally’ read primarily a familiarity with American mainstream and popular spoken and written language from the latter half of the twentieth century, of a kind fairly reflected in newspapers, TV programmes, and everyday conversations? Or would a corpus of literary fiction by Carver’s North American near contemporaries – e.g., Bellow, Updike, Coover, Didion, and Barth – be a better comparative background? These speculations have no obvious resolution. Where too large a cultural or historical gulf opens up between the target text and the comparator corpus (e.g., a reference corpus of Early Modern English, with a Carver story as target text), the keyword and other findings will be increasingly unreliable.

Still, some variables that may loom large or dwindle into the background, depending on corpus-selection, would seem to be worth taking particular care over. One of these relates to the tense of narration: one story I have been working on recently is Ann Beattie’s “Cosmos”, and somewhat to my consternation I found that relative to my first reference corpus of twentieth-century fiction, the top keyword in “Cosmos” was says (with a name, Carl, as second topmost keyword only). But this result stemmed from the fact that “Cosmos” uses the historic present for narration, while nearly all the work in that first comparator corpus, nearly all of which was written pre-1970, used the more traditional past tense. As a result, said was frequent in the reference corpus, but says was not. Nor is says especially frequent in “Cosmos”: it is 0.93% of the text, while another shorter Beattie story also in the historic present, “Sweetest Song”, has a textual frequency for says of just over 2%. So the seeming ‘prominence’ of says in a text (against a reference corpus of earlier modern fiction) or said (against a corpus of contemporary fiction) can easily be an artefact of the test procedure using an ill-matched reference corpus.

At the same time, items like says (in stories in the historic present) or said (in stories told in the preterite) are also a reminder of how the traditional lexical-grammatical division, with the temptation to set aside or discount the distribution of grammatical items as ‘mere framework’, may blind us to some complexities.
Corpus analysts urge us to avoid pre-judging words as lexical or grammatical, semantic or structuring, and instead to entertain the possibility that most ‘grammatical’ items have some semantic and lexical force, and – an implication of the Firth/Sinclair idiom principle or phraseological tendency – that many lexical items have some syntagmatic force, some tendency to collocate or colligate with certain lexicalised partners and not others. In vacuo, one would without hesitation classify *says* and *said* as lexical items; but in the context of literary narratives, so often involving reported speech, and so systematically narrated, normally, in the past tense or, currently, in the present tense, *says* and *said* seem to have some of the structuring qualities of grammatical items. They are not ‘closed class’ items, but they are so much the default verbs of communication in the reporting of direct speech that they are like a colourless prototype, against which more ‘lexical’ and informative variants are in contrast.

4. A note on the stories selected for analysis

The stories on which my analyses focus include Joyce’s “Two Gallants”, Mansfield’s “Bliss”, Raymond Carver’s “Cathedral”, “Boxes”, and “A Small Good Thing”, Updike’s “A& P”, and Alice Munro’s “The Love of a Good Woman”. None of these is part of the reference corpus used for comparative calculations: the latter is a corpus of my own compilation, comprising approximately 500,000 words, and containing 20th century novel and short story fiction by such British or North American authors as Joyce, Hemingway, Faulkner, Woolf, Mansfield, Updike, Bellow, Bowen, Trevor, Carver, Beattie, Atwood, and Munro. Only in some senses of the term can either the stories or the reference corpus be characterised as ‘representative’, and like any corpus my reference one may be found lacking in one respect or another. Even with the usual caveats, a literary story like “Barn Burning” is unique, exceptional in its particularities of phrasing as well as when viewed as a whole, in ways that, at the very least, are much less pronounced in individual non-literary texts. And even the corpus can be no more than a sample of the range of most frequently-taught and -studied 20th century short fiction. But I will argue that there are family resemblances among the stories analysed, and that even with their diverse patterns of grammar, phrasing, characterisation, plotting and texture, they are representative of the works that have informed a modern reader’s ‘competence’ with regard to narrative progression in the modern short story.
5. Conclusion

In this book I examine lexical patterns of relation and inferable progression across much longer ranges than adjacent sentences, or all the sentences in a paragraph. A vivid example of truly long-distance repetition was the lexical bond which Hoey identified between one sentence in a book and two later ones which were approximately 4,000 sentences or 260 pages distant (Hoey 1991: 151). I am looking at the uncharted territory that is an entire short narrative. Are there kinds of connectedness, constrainings of the domain from which the reader can reasonably expect crucial subsequent text to come, created by earlier-occurring text in a short story? In an intuitive or informal way, most people are quite ready to affirm that story openings and continuations create a setting and a mood, and ‘launch’ an arc or trajectory of events such that a general sense of the likely endpoint, the outcome, forms in the reader’s mind. Nearly everyone will assent to these ideas. It is not that the reader knows for sure what will happen next, or finally, or with what consequences for which participants, and with what intermediate and final impressions on that reader. It is rather that the reader “can’t help” forming some ideas, some judgements, on these matters. And for some readers this is no more than a ‘feeling’, which they cannot put into words--while other readers, suitably prompted, may be able to comment at length on their expectations. Where do these expectations come from? I argue that to an important degree, in creative short fiction that rises above the level of mechanical reproduction of generic form and content, these expectations are ‘in’ the text; their source is the particular lexical-discursive composition of that part of the story which the reader has read so far, where to read is of course to understand, to absorb, to encapsulate, and to engage with intellectually and emotionally.

This book’s exploring of narrative progression applies and tests corpus linguistic methods in particular, these being ones that have begun to be extensively adopted in some fields of stylistics. As briefly noted earlier, an enormous amount of work has been done on the computational or semi-automated identification of text structure and coherence, and textual reception and interpretation, in the fields of empirical study of literature, cognitive psychology, and computational linguistics; and these fields frequently draw on corpora of texts, to test and refine their procedures. Among many competing analytical instruments one might mention, for example, Coh-metrix and Latent Semantic Analysis. Coh-metrix is a web-based software tool that can measure aspects of a text’s discourse cohesion and language complexity. The textual features it can analyse include characteristics of words, syntax, referential cohesion, semantic cohesion, and aspects of the inferable situation (see Graesser, McNamara, Louwerse & Cai 2004, and Graesser, Jeon, Cai, and McNamara. 2008, for fuller description); the software has proven
useful in the study of text coherence and readability, with wider applications now emerging. Latent Semantic Analysis (LSA) is used to identify the theme of a text and create a summary by automatic computational means (Landauer, Foltz and Latham 1998; see also Kintsch 2002). It does this by representing the meanings of the words in a text as a multi-dimensional (300 dimensions) semantic vector, with similarity of different words’ vectors correlating with their semantic closeness. By this means LSA can automate the selection of the most typical and most important sentences in each section of a text, or the whole text.

The present corpus stylistic project remains distinct from the more psychologically-oriented and often computationally-framed traditions of research of which Co-Metrix and LSA and Wordnet (Felbaum 1998) are representative; but the latter are more directly relevant and therefore receive more attention in a separate study (Toolan, forthcoming) which is in some respects complementary to this one. That separate study begins with Hoey (1991), and explores in depth the nature of the textual resources that contribute to ‘immersion,’ the emotional and ethical involvement which is arguably a hallmark of literary reading (see Louwerse and Kuiken 2004; Miall 2006).

In these opening two chapters I have tried to make the case for looking in detail – stylistic and corpus analytic detail – at the ways short story narratives create expectations in readers. Everything in a literary text, all the choices of wording and structure and sequence, contribute to the creation of expectations. But it is axiomatic for the present enquiry that some things are more important than others, and that there is an organisation of information and effect in these stories that rewards close study. Why is expectation important? Because it is directly reflective of our engaging with literature as an emotional and cognitive experience (as was long ago emphasised by Rosenblatt 1938). We turn to literature not only because it makes us think but because it makes us feel, and for many of us the ‘total physical response’ thinking and feeling we experience during the reading of literature regularly surpasses the imaginative and emotional insight we derive from other valuable kinds of writing – history, science, psychology, sociology, theology, biography – and even from other kinds of art. This is a study of the particular textual means by which expectations are created, on the assumption that those expectations directly give rise to responses of confidence and then surprise, suspense and then usually the release of suspense, anxiety and then its resolution, and similar experienced sequences. In a sense expectations are both thoughts and emotions, so as a theoretical concept they somewhat resist the questionable division of textual information and effects into an ideational part and an emotional part; they emphasise the holism in the text and in the reader’s developing apprehension of the text.
CHAPTER 3

Lexical patternings in short stories

1 Word frequencies in “Two Gallants”

Corpus linguistics can uncover some of the patterns of lexical emphasis and forms of foregrounded repetition in a text. Can these help us to pinpoint the core or key articulation points of the experienced narrative sequence? The goal here is to use corpus methods to identify elements in the stories’ textual progression and prospection, treating these as key sources of complex reader expectations and reactions, both cognitive and emotional. In this and the following chapters I present some of the results I have obtained, adopting procedures influenced by the work of some of the corpus linguists discussed in Chapter 2, but also mindful of the difficulties encountered by a links-and-bonds approach to lexical patterning in narratives (see Toolan, forthcoming, for a fuller discussion of these difficulties). To help in identifying textual patterning via lexical repetition, I also use the WordSmith Tools Keywords procedure, which identifies the disproportionately frequent words in a text. The listing that Keywords produces is comparative: using as norms the frequency of occurrence of word-forms in some chosen appropriate reference corpus of texts, Keywords uses a statistical calculation to identify and rank those forms in the analysed text that are markedly more (or less) frequent than the reference corpus frequency norms would lead us to expect. The story I will chiefly focus on in this and the following chapter is Joyce’s “Two Gallants”.

1.1 “Gallantry” in Dubliners

Corpus analytic methods can help us to identify at least the following features of a text: its keywords; its strange collocations (including its flouting or breaching of idioms); its disparities of topic and wording; and its intertextual lexical echoes and reiterations of other texts. In the case of a short story in a linked collection of stories, an obvious first intertextual linkage to explore is that between individual stories and others in the collection. Initially, this may seem little different from what has long been attempted in one branch of literary criticism, that is, the studying of the distributed recurrence of words and imagery in plays and novels, as a foundation for thematic interpretation. In English literary studies, for example,
Shakespeare’s plays were unsurprisingly among the first texts to attract such detailed scrutiny in the studies of patterns of dominant words and imagery by Kolbe (1930) and Clemen (1951). But the technology-enabled systematics and rapidity of the corpus approach soon altered the nature of the analysis.

An appropriate place to embark on the lexical analysis of James Joyce’s “Two Gallants” is with that archaic word gallants in the title, archaic even at Joyce’s time of writing, circa 1910. Where else is it, or variants of the lemma, used in *Dubliners*? A simple search rapidly finds five occurrences beyond the story title, in which a word containing the stem gallant, with or without preceding or following morphemes, is used. Below, I follow the convention of placing the stem between asterisks—*gal*—to denote this range of possible forms: gallants, ungallant, gallantry, gallantly, etc. The five non-titular instances in *Dubliners* occur in “After the Race”:

*The journey laid a magical finger on the genuine pulse of life and gallantly the machinery of human nerves strove to answer the bounding courses of the swift blue animal.*

in “Two Gallants”:

*In his imagination he beheld the pair of lovers walking along some dark road; he heard Corley’s voice in deep energetic gallantries and saw again the leer of the young woman’s mouth.*

in “Grace”:

*In her [Mrs Kernan’s] days of courtship, Mr. Kernan had seemed to her a not ungallant figure*

and twice in “The Dead”, first when Mr Browne brings Aunt Julia forward to sing:

*Mr. Browne was advancing from the door, gallantly escorting Aunt Julia, who leaned upon his arm, smiling and hanging her head.*

and then when Gabriel is about to propose a toast to the Three Graces of Dublin:

*He raised his glass of port gallantly.*

These six uses of *gallant* are a simple kind of intertextuality, binding the stories of the collection together, but the fact is that it would be immensely laborious to do such a search without an electronic file and simple search mechanisms – and to do it not just for the word gallants from this story title, but also for other semantically rich words in other titles in the *Dubliners* collection: counterparts, sisters, dead, painful, grace, and so on – on the assumption that titles are in a sense inescapably key words. As the six instances in *Dubliners* reflect, the uses of *gallant* in various word-classes (noun, adjective, and adverb) are gendered and highly
evaluative. They are only used of men, who have or affect a gentlemanly or considerate if patronising manner towards women. In all uses here, not merely in the clearly ironic application to Corley and Lenehan, there is a counterfactual intimation, a conscious going beyond what is entirely reasonable or justified, in the action described as gallant. There is little true gallantry in Mr Browne escorting Aunt Julia, or in Gabriel raising his glass of port, in all the given circumstances; or it is a ‘gallantry’ far removed from war-hero bravery. Several of them hint at the narrator’s irony or the character’s insincerity, or both. In this they are somewhat similar to what Louw has proposed regarding the ‘semantic prosody’ of certain common phrases (described in Chapter 2). He suggests that when we encounter unnatural semantic prosodies (what Mary did caused a lot of happiness) we suspect either speaker irony or speaker insincerity (Louw 1993).

1.2 High and low frequency lexis in “Two Gallants”

Relatively simple corpus analysis (using a package such as WordSmith Tools) reveals some details not so apparent to the ordinary reader. For example, using Wordlist to produce a concordance reveals that the names Corley and Lenehan are among the top twenty most frequent words in the story, but that Lenehan ranks 17th, with 29 occurrences, while Corley is 11th, with as many as 46 occurrences. A moment’s reflection can explain this initially surprising trend: Lenehan, the focaliser and often the focalised, is extensively denoted by personal pronouns rather than name. Roughly the reverse is the case with Corley. A more interesting piece of evidence may be the fact that after the occurrence of some of the more predictable grammatical items and character names as the story’s most frequent words – the, he, of, and, Corley, etc. – the most frequent lexical items in the story are the following: said (45) street (26) young (14) face (13) know (12), walked (12), and eyes (11). Said is frequent in its use in the reporting of speech (see the remarks on said and says in past and present tense narration, in Chapter 2, Section 3 ); street will be returned to below.

A more detailed study can then follow – for example, examining the fourteen instances of young in their co-textual settings, i.e., examining the concordance lines for young, which WordSmith’s Wordlist program retrieves automatically. These lines reveal that all fourteen occurrences of young precede either men (4, always itself preceded by two: (the) two young men), man (1), or, and predominantly, woman (9). And of the twelve instances of walked found, only one does not refer to Lenehan: this otherwise lexically-diverse text is thus quite emphatically and purposefully repetitive in its telling that Lenehan walked and walked. Lenehan is linked to most of the eleven noted mentions in the text of eyes, too: eight of
them refer to Lenehan’s eyes, in situations where they are quasi-agentive, such as *His eyes searched the street* (see discussion in Nørgaard 2003; and Toolan 2007).

Similarly, closer scrutiny of the contexts of use of the high-frequency word *face* suggests how central it is to the core characterisation and concomitant understanding of Lenehan that the narrative projects. Consider this intrinsically inconsequential moment in the middle of the story, the doldrums during which Lenehan is killing time as best he can with limited funds, until Corley returns. Lenehan has entered a cheap restaurant:

> He spoke roughly in order to belie his air of gentility for his entry had been followed by a pause of talk. His face was heated. To appear natural he pushed his cap back on his head and planted his elbows on the table.

It is his appearance, his air of gentility, and the assessment others make of him that we infer is at the forefront of Lenehan’s mind. These prompt his chameleon attempts at ‘passing’ (his speaking roughly and trying to appear natural) rather than any concern over an inner, challenged identity. If in the story’s opening description of Lenehan the phrasing was ambiguous as to whether there was something dissembling or presentational about how the yachting cap “was shoved far back from his forehead”, here the artificial element or contrivance is explicit: he pushed the cap back “to appear natural”.

One word is notable for its absence from this story, by comparison with every one of the other *Dubliners* stories, namely the form of address *Mr*. There are no designated *Mr*’s whatsoever in this story (only two gallants). The only other key-words uncovered are personal names (*Corley, Lenehan, Corley’s*), a common enough tendency in literary narratives – but not one that should be simply recorded: it could be revealing to find literary narratives where personal names were not key, and consider the implications. I give the above brief examples of kinds of word-frequency evidence to suggest the possibilities of simple frequency and collocation analysis in narrative analysis, rather than to embark on a lengthy exploration at this stage.

2. **Textual richness measured by lexical diversity**

2.1 The type-token measure

Simple corpus analysis can also help confirm textual richness. A basic (but not uniformly reliable) measure of such richness is the comparative diversity of different words in a text, known as the type-token ratio. If text A of 4,000 words (tokens) contains 1,000 different words (types), then that text is deemed lexically more
diverse and less repetitive, arguably lexically ‘richer’, than text B of the same length but with only 500 different words or types. This can also be referred to as “lexical density”. Text A would be said to have a type/token ratio of 25 (effectively a percentage), meaning also that on average each word in the text is used four times; while text B has the lower type/token ratio of 12.5 (on average each word in text B is used eight times).

Type/token scores only scratch the surface of textural richness, of course, in no way addressing the figurativeness of the language, or the semantic or collocational associatedness of the words that occur, or the syntactic complexity or simplicity of sentences, or complexities and ambiguities of reference. In addition, the type-token scores of very short texts, not surprisingly, are usually quite high and an unreliable indicator; normally the ratio declines for any text, as more of its gradually more repetitive extent is included in the calculation.

Despite all those caveats, it can be helpful to compare the type/token ratio for the whole of ‘Two Gallants’ (10.4), with the type/token ratio of the second paragraph of the story, in which the main characters of the story are first introduced (the paragraph is supplied below). This single paragraph has a type/token ratio that is six times higher than that for the entire story (62, as against 10.4): its 205 word tokens involve 127 different word types. In fact, what repetition there is is almost entirely of grammatical items – multiple uses of he, the, his, was, and, of, and so on – with very few re-used lexical words or phrases: companion, face, waves of expression, slung, waterproof. Other than in those very few cases of grammatical repetition, lexically the paragraph of the text is creating a continual diversification, which calls for an answering interpretive expansiveness in the reader. In the penultimate sentence, for example, readers must incorporate (new) ideas of breeches, whiteness, rubber, shoes, and several more items, in their construal of the story.

A paragraph or passage with a very high type-token ratio as this one has is in some respects at the opposite end of the spectrum from a passage in which there are multiple intersentential repetition links (to use Hoey’s [1991] terminology). A clear example of this contrasting intersententially-bonded texture is evident in the story’s short opening paragraph, which of course immediately precedes the one just discussed. Both paragraphs are reproduced below. The expression grey warm evening in the first sentence immediately generates a triple-repetition bond with the appearance of the same three words in the final sentence, in the phrase warm grey evening. Those two sentences have other repetition links with each other and with the intervening sentences besides; the least contentious of these are shown below in bold. Like Hoey 1991 I classify, e.g., coloured and hue as a repetition pair, despite the morphological difference, on the grounds of the frequent intersubstitutability of colour and hue. More debateable would be, for example, to treat changing unceasingly as repetitively linked with circulated; so this is not marked. Nor are
merely intrasentential repetitions (e.g. *changing* and *unchanging*, *unceasingly* and *unceasing*, in the final sentence of this first paragraph).

THE grey warm evening of August had descended upon the city and a mild warm air, a memory of summer, circulated in the streets. The streets, shuttered for the repose of Sunday, swarmed with a gaily coloured crowd. Like illumined pearls the lamps shone from the summits of their tall poles upon the living texture below which, changing shape and hue unceasingly, sent up into the warm grey evening air an unchanging unceasing murmur.

Two young men came down the hill of Rutland Square. One of them was just bringing a long monologue to a close. The other, who walked on the verge of the path and was at times obliged to step on to the road, owing to his companion’s rudeness, wore an amused listening face. He was squat and ruddy. A yachting cap was shoved far back from his forehead and the narrative to which he listened made constant waves of expression break forth over his face from the corners of his nose and eyes and mouth. Little jets of wheezing laughter followed one another out of his convulsed body. His eyes, twinkling with cunning enjoyment, glanced at every moment towards his companion’s face. Once or twice he rearranged the light waterproof which he had slung over one shoulder in toreador fashion. His breeches, his white rubber shoes and his jauntily slung waterproof expressed youth. But his figure fell into rotundity at the waist, his hair was scant and grey and his face, when the waves of expression had passed over it, had a ravaged look.

Thus the opening two paragraphs of “Two Gallants” comprise a highly repetitively bonded paragraph followed by a highly unrepitative bond-less one. This is not to say that there is no cohesion or patterning in the second paragraph: there is, but it chiefly depends from the recurrent pronominal references to “The other” [of the young men], first particularised in its third sentence and thereafter referred to by *he* or *his*. This means that, recurrently, at least one (and usually only one) grammatical cohesive link is present between later sentences and earlier ones in this paragraph. For example, in the sentence *Little jets of wheezing laughter followed one another out of his convulsed body*, only the *his* entails a link with the antecedent *The other [of the two young men]* in sentence (3). And the same could be said of all of the sentences after (3), namely that each is linked, albeit only along the single chain begun with *The other*. Thus what the paragraph lacks in the way of multi-link intersentential bonds, it arguably makes up for in the persistence of its one cohesive co-referential chain. That chain not only persists. It is also structurally prominent: as the paragraph proceeds, a *he* or *his* is increasingly often found in sentential Subject or Theme position. In fairness, there are some instances of lexical repetition (*face, eyes, waves of expression, companion*) and semantic-field association (*amused, laughter, enjoyment*) in the second paragraph. But the structural
unity of this paragraph is grounded in the grammatical reference cohesion (which enables there to be great lexical diversity) while in the first paragraph the integrity comes from the lexical bonds, not grammatical ones.

2.2 Lexical innovation and text structure

The previous section has again approached the subject of the tension, found in all texts but especially relevant to narrative text, between repetition and change, telling the same thing again versus telling something different but related. Can that abstract and general impression be connected to further textual evidence, the better to understand narrative textual progression? We have seen that the first paragraph of “Two Gallants”, taken on its own, is highly repetitive while the second paragraph, again taken separately, is highly lexically diverse; with greater effort one could show that the second paragraph additionally does not draw on the lexis of the first paragraph, so that it has very few lexical links with the previous paragraph and is therefore all the more novel to the reader. What about the third paragraph, and its lexical innovativeness (or otherwise) relative to previous paragraphs, and the fourth, and so on? We need reliable automated procedures by which some version of cumulative repetitiveness, across a text, and some kind of non-repetitive diversity or innovation, can be measured.

Various programs are now available that, with a high degree of accuracy, will measure lexical innovation along the length of a text. One such program is Paraword (Mason, n.d; see http://phrasys.net/service/paraword.html), which exploits the type/token principle and calculates the degree of lexical novelty that each successive segment of a text displays, relative to the stack of segments that precedes it. It is for the analyst to decide how to sub-divide the text into successive segments that, it may be hoped, will reflect its progression (e.g., treating each paragraph as a segment, or each printed page, or each chapter, or, if applicable, each section implicitly designated by the author as distinct by interposed blank lines on the page). In the case of “Two Gallants”, I divided the text into 13 sections, of varying length but with a 300-word average, largely on the basis of the story’s episodic shifts (cf Emmott’s [1997] contextual frames, or what Werth [1999] called ‘incrementations’). Some sections are quite short. For example, I treated each of the story’s first two paragraphs as complete sections. The opening words of the 13 sections I have proposed are, in order, as follows:

1. The grey warm evening of August had descended…
2. Two young men came down the hill of Rutland Square.
3. When he was quite sure that the narrative had ended…
4. Corley’s stride acknowledged the compliment.
5. --Well... tell me, Corley, I suppose you’ll be able
6. They walked along Nassau Street and then turned into Kildare Street.
7. --There she is! said Corley.
8. Corley did not answer. He sauntered across the road…
9. Now that he was alone his face looked older.
10. He was hungry for, except some biscuits…
11. He paid twopence halfpenny to the slatternly girl and went out…
12. His mind became active again.
13. Lenehan hurried on in the same direction.

Paraword calculates the percentage of new words in each segment, and can then either display (and print) the complete text with the new words highlighted, or list the new words only. The general idea and expectation is that in a text of 13 roughly-equal-length sections all on a unified topic, there will be a steady decline in the percentage of new words introduced, with each succeeding section. Here now are the actual percentages of new words that Paraword finds in each successive section:

71, 56, 42, 36, 22, 28, 27, 24, 17, 15, 18

The most noticeable departure from a fairly steady decline in lexical novelty, as the story extends, comes in the sixth section, which is preceded and followed by sections with novelty percentages of 26 and 22 respectively but has the higher frequency of 35. The text that I had designated as the sixth section in the story is the brief encounter with the weary harpist, an intuitively highly distinct segment in the story.

Somewhat similar to Paraword is Youmans’s Vocabulary Management Profile (VMP) software, which produces a graph of the variability of a text’s lexis, plotting frequency of lexical novelty on the vertical y-axis against successive textual segments (which must be of a consistent length: e.g. 55, or 100, or 300 words) on the horizontal x-axis. Where a ‘spike’ in the graph occurs, away from the expected steady decline in lexical novelty, something significant may be occurring: a spike downwards (exceptionally high lexical recycling) may reflect a discourse that is ‘becalmed’, while a spike upwards may reflect major new developments or abrupt change in the discourse. The VMP software allows for several distinct but related tests, and different tests and segment spans produce slightly different pictures; trial and error may help identify the most useful combination of test and span. Using version 2.2 of the VMP on the story, 2nd pass through, with intervals of 201 words, reveals peaks in the gradually declining graph of word novelty at around word 1,650 in the text, 2,100 (the main peak), 2,550, and a smaller rise at word 2,900, with noticeable low dips at words 580 and 3,125. Those approximate readings are from the graphic representation. More precise identifications can be made by using the detailed statistics VMP also provides, on the average moving type-token ratios (i.e. in the span extending 100 words either side of any given word).
As noted, there are two ‘dips’ in the VMP graph, where relatively old or recycled vocabulary is identified, sites of topic-ebb in contrast to topic-flow. One is early in the story, and the other is around word 3,200, which comes in the sentence *The crowd of girls and young men had thinned and on his way up the street he heard many groups and couples bidding one another good-night.* Arguably the entire paragraph in which this sentence occurs is a relatively ‘flat’ one, low in narrativity, preparatory only to the more dramatic and dynamic flow of actions and thoughts and reactions from Lenehan, full of suspense, with which the story ends:

He left his friends at a quarter to ten and went up George's Street. He turned to the left at the City Markets and walked on into Grafton Street. The crowd of girls and young men had thinned and on his way up the street he heard many groups and couples bidding one another good-night. He went as far as the clock of the College of Surgeons: it was on the stroke of ten. He set off briskly along the northern side of the Green hurrying for fear Corley should return too soon. When he reached the corner of Merrion Street he took his stand in the shadow of a lamp and brought out one of the cigarettes which he had reserved and lit it. He leaned against the lamp-post and kept his gaze fixed on the part from which he expected to see Corley and the young woman return.

The paragraph is one of preparation and not-yet-tense waiting; nothing of any particular narrative interest happens within it. It might be argued that it is stylistically designed by Joyce to denote the passing or biding of time, while the denouement draws near.

Concerning the main upward ‘spike’ already noticed in the area of token 2,100, it emerges that the average type-token ratios rise to scores of 0.35 and above with word #2,129, red, in *red flowers was pinned* and stay above 0.35 until word #2,329, *turned in he turned about and went back.* So the upward peak, its topmost point being word #2,242, *waiting* (emboldened below) is the following passage. Around the peak of the spike, I have underlined all those words where the type-token ratio is at 0.4 or above.

The ends of her tulle collarette had been carefully disordered and a big bunch of red flowers was pinned in her bosom stems upwards. Lenehan’s eyes noted approvingly her stout short muscular body. Frank rude health glowed in her face, on her fat red cheeks and in her unabashed blue eyes. Her features were blunt. She had broad nostrils, a straggling mouth which lay open in a contented leer, and two projecting front teeth. As he passed Lenehan took off his cap and, after about ten seconds, Corley returned a salute to the air. This he did by raising his hand vaguely and pensively changing the angle of position of his hat.

Lenehan walked as far as the Shelbourne Hotel where he halted and waited. *After waiting* for a little time he saw them coming towards him and, when they turned to the right, he followed them, stepping lightly in his white shoes, down
one side of Merrion Square. As he walked on slowly, timing his pace to theirs, he watched Corley’s head which turned at every moment towards the young woman’s face like a big ball revolving on a pivot. He kept the pair in view until he had seen them climbing the stairs of the Donnybrook tram; then he turned about and went back the way he had come.

Why should this be the most relatively lexically new, new-topic-introducing, passage in the story? It may be because it is the only point where all three main figures are co-present, with the text topic switching relatively rapidly from the woman, to Corley, to Corley and the woman, rather than dwelling, as it mostly does, on Lenehan. But then the reader may wonder why VMP did not especially identify the short passage about the weary harpist as most novel (as Paraword did). The explanation appears to be that the latter passage, which begins around word 1,600, is not extensive enough to sustain a raised novelty-rating when a 201-word search-span is used (even though VMP, as indicated, records a secondary peak at 1,650). And this in turn implies that to some extent automatic identification of lexical innovation is a function of the size of text-span selected in the calculations. But whatever the mechanism adopted for tracking vocabulary innovation through a text, the scores that emerge appear unsuited to pinpointing the particularities of narrativity or prospection, and more useful in the identification of broad textual segmentation.

3. Foregrounding via repetitive phrasing
   or by novelty of phrasing and collocation

In this section I will mention very briefly one mechanical means of identifying the repeated multi-word sequences in a text (not always ‘phrases’ in the grammatical sense), since such repetitions of verbal ‘chunks’ may be quite influential in the impression a story makes on the reader. Then I will discuss at greater length a sharply contrasting phenomenon: the textually (and even intertextually) unique phrasing, whose novelty seems importantly advertised in its context of occurrence.

3.1 Clusters

Corpus linguists who are interested in multi-word repetitions avoid the term ‘phrase’, because of its syntactic implications. In this section is a phrase – prepositional – while this section I is not; but the sequence this section I could in principle be found repeated more often than In this section in a particular text. The corpus linguistic label for identical multi-word chunks or sequences like this section I is clusters or lexical bundles or n-grams (where ‘n’ represents any reasonable number
of adjacent words, such as 3, 4, 5, etc.). Strictly speaking a n-word cluster is any occurring sequence of n words. So the previous sentence comprises the following series of three-word clusters: **Strictly speaking a, speaking a n-word, a n-word cluster, n-word cluster is**, and so on. This is usually uninteresting of course, and attention turns rather to clusters that are used at least twice in a text and therefore arguably comparatively noticeable. Dr Martin Luther King's civil rights speech in Washington D.C. in 1963 unforgettably used the four-word cluster *I have a dream* many times over.

The repeated n-grams or clusters of a text can be rapidly uncovered with a variety of available programs, including WordSmith Tools (Scott 2004), and Collocate (Barlow 2004), another computer program. The idea is an extension of traditional textual criticism, where the critic might have a hunch that a particular phrasing is noticeably frequent in a novel or play, but these programs make no silent pre-selection, since they simply list and count all multi-word sequences. The programs naturally report different results, depending on the settings one selects. If the search asks for all the three-word sequences that occur at least three times in a long text, this might generate an extensive list of sequences, many of which might appear quite uninteresting. Or you could seek, from the same text, just those five-word sequences that occur at least six times, and find that there are very few (perhaps too few to be revealing). In short there is no single template that will guarantee that informative n-grams or clusters will be extracted from a text, and perhaps this is particularly so when literary or fictional texts are explored.

Just which clusters to attend to needs assessment on a case by case basis. In a text of just 4,000 words like “Two Gallants”, there is little scope for high-frequency multi-word clusters and indeed clusters may be of minor importance, whereas in a novel one's interest might focus on comparatively frequent four-word clusters (to the disregard of much less frequent six-word clusters). Between these extremes, a 25,000-word story like Alice Munro's “The Love of a Good Woman” is long enough for clusters to be of major potential interest. In “Two Gallants” however, while there are 18 three-word clusters that occur at least three times (including the corner of and the young woman), only one four-word cluster occurs more than twice, and this is *it off all right*, which appears in the following dispersed sentences, where it refers to the vaguely-specified crux of what plot there is to the story:

> Well... tell me, Corley, I suppose you'll be able to pull it off all right, eh?  
> But tell me, said Lenehan again, are you sure you can bring it off all right?  
> But the memory of Corley's slowly revolving head calmed him somewhat: he was sure Corley would pull it off all right.

Often the corpus linguistic interest in clusters is because they are typical or indicative of particular genres and text types, but in literary analysis they are more likely
to be of interest if they are prominent just in single texts. It may be that, in novels especially, there are both clusters and “key clusters” (just as there are frequent words and keywords), the key clusters of a text being the ones that are comparatively disproportionately frequent, but not necessarily the most frequent ones. Unfortunately, so far as I am aware, there is no widely-available software at present for identifying key clusters. Relative to clusters in general, text-specific key clusters would be more likely to evoke a specific set of associations or ideas relevant to that narrative text. And what they evoke may tend to be more semantically focussed and delimited than readers’ associations with shorter segments such as single words in the text. This reflects the broad assumption (other things being equal, and allowing for exceptions) that the longer the chunk of text, the more likely that, relative to one- and two-word segments that surround and predominate, the chunk will evoke in the reader a mental picturing.

These general ideas can be applied to *it off all right* if we take this to be a – perhaps the – key cluster in “Two Gallants”. Like the outer borders of the lexical unit (Sinclair 2004), the key cluster has semantic or colligational tendencies at its “edges”. Here, to the left of *it off all right*, we expect and find a verb such as *bring* or *pull*, preceded by the naming of an individual as Agent (e.g., *you*, *him*, or *Corley*), and also some modal of ability or intention. The whole evokes a mental picturing of a single risky enterprise or goal, difficult to achieve, where calculation and luck are involved. It is only in context, as we gradually realise that the phrase is used by the entirely ungallant Lenehan to describe the dishonest receiving of money from Corley’s woman (or her employer), that its distasteful use here is fully understood. The bringing of something scarcely nameable off all right lies at the core of the story’s narrative drive; and the analysis of the story’s lexical clusters reveals that the kernel of the phrasing is also exceptionally prominent in the text’s wording.

3.2 Achieving the stern task of living: a tissue of not-so-vague associations

Of Lenehan we are told, near the opening of the story:

> No one knew how he achieved the stern task of living, but his name was vaguely associated with racing tissues.

Anyone who admires a finely crafted sentence, fermenting with irony, will savour this. Literary criticism will tempt us to comment on its use of indirectness and relexicalisation, its Bakhtinian (or pseudo-Bakhtinian) ‘clash’ of a rather puritanical perspective on life (it is a ‘stern task’) and Lenehan’s hinted-at disreputable *modus vivendi* (contributing to newspapers devoted to horseracing and gambling). As reader-critics we may suspect that we are intended to recognise that the narrator intends us to find *achieved the stern task of living* an awkward and revealing
phrase, obliquely critical of Lenehan. But can these impressions be bolstered by corpus analysis: does corpus evidence confirm that the narrator is conveying this complex and subtle evaluation?

One way of demonstrating the interestingness of this formulation is to show that it involves a series of rare to non-existent incorporated phrasings. That is to say, not only is ‘achieved the stern task of living’ a unique formulation, perhaps never previously used in writing, but this larger sequence’s component phrasings are strikingly rare in other corpora. This claim can be supported by checking the frequency of those component phrasings in a really large corpus (such as the Cobuild Bank of English), or via a web search engine such as Google. Below are frequencies of various sub-parts of the full phrase (achieved the stern task of living) as returned by Google when a search was made on December 29, 2006:

- the task of living 21,600 hits
- achieved the task 22,700 hits
- achieved the stern task 66 (all and only from web versions of “Two Gallants”)
- the stern task 224 only!
- the stern task of living 63 (all but one of these was from the story text, or was referring to the story; the single exception puts the phrase in quotes and, one may surmise, comes from a writer aware of the Joycean original)

These figures need further comment. In particular, the task of living (as a complete phrase) and achieved the task (as one phrase) were far less frequent than the somewhat indiscriminate retrieval of over twenty thousand ‘instances’ by Google would suggest. Thus there were few instances of the task of living used in just the way it occurs in the story, with living occurring at the phrase boundary; most found instances involved some form of postmodification of living (so that the phrase does not refer to making a living in general, but to living under certain conditions: the task of living on campus, the task of living with central vision loss). Somewhat similarly, the Google-retrieved instances of achieved the task include a majority of instances in which achieved and the task occur in separate syntactic clauses, separated by a comma or period. By contrast the sequence accomplished the task is much more frequent, with 98,300 hits, and overwhelmingly these three words are phrase partners in those instances.

As for the 224 uses of the stern task, these include a number of links to electronic versions of the Joyce story, unsurprisingly; also one finds that at least one academic has used the phrase in the title of an essay about Dubliners (Wallace 2001). A good example of the intertextual borrowing of part of the phrase by a later writer
Narrative Progression in the Short Story

occurs in the final sentence of a commentary by Professor Zack Bowen, writing about what is involved in serving as Chair of a university English department:

It is our job to try to replace righteousness with understanding...and to ensure that we and our colleagues get on with the stern task of surviving by enjoying the whole enterprise as much as we can. (Bowen, 1997: 9, emphasis added)

To summarise, then, frequency scores from a large general corpus or textbank (such as Google, or the Bank of English) can help confirm that achieved the task is markedly infrequent by comparison with someone having accomplished or completed or pursued a task; that the task of living (as a complete phrase) is strikingly rare by comparison with the task of living + complement; and that achieved the stern task of living is so marked that it is simply unattested, outside the unique context of the Joyce story.

One way to use the corpus data to sharpen one’s intuitions about the source of the sentence’s specialness is to ask: How few words from the target phrase or sentence need one take, in coherent phrasal combination, before finding that all search-engine instances are from the unique, source text? The frequencies listed above point to the four-word phrase achieved the stern task as being the shortest unique phrasal sequence in the sentence (just surpassing the almost equally unique – but for one use – five-word string, the stern task of living).

In other words, as soon as you type “achieved the stern task” into the searchbar of Google, you can be sure that “Two Gallants” is the only text you will be directed to, i.e., that “Two Gallants” is the only text Google knows of containing these four words in this sequence. This was most recently checked by me on 20 May 2008: all 13 hits for the phrase were from versions of the story. Of course if a version of the present discussion should be loaded onto some website in the future…! By contrast no phrasally-complete three-word sequence here, such as the stern task is, according to Google, comparably unique to “Two Gallants”; and while nearly all instances of achieved the stern are from the Joyce story, the three-word sequence is an incomplete phrase, and the Google hits also include some cross-phrasal instances (such as …achieved. The stern…).

We may wonder why, in a sentence with multiple mildly grotesque and ironic formulations, achieved the stern task is so especially suspect. A corpus linguistic answer draws us to ideas about what has been termed the semantic or discourse prosody of a phrase’s recurrent patterns of use (Sinclair 2004; Louw 1993). A related strong but controversial claim, introduced in the previous chapter, is that of Louw (1993), who suggests that where a word or sequence is used in an unusual way, at odds with its usual prosody, then this necessarily implies either speaker irony or speaker insincerity. So if utterly normally patterns with negative evaluations
(utterly devoid of redeeming qualities), then a remark like The garden was utterly full of varied colours and textures should not be taken at face value.

In the present case we might suggest that achieved is strongly primed to collocate with successes and positive things, success or completion after effort or despite difficulties, so that stern task is a highly unusual collocate (since it is explicitly onerous, a chore, a duty rather than a pleasure). These taken together, (how he) achieved the stern task of, arguably prime the reader for some punitive or burdensome activity as phrasal complement (such as paying his taxes, or providing for his many dependents); but again expectations are flouted by the word living, which is not in itself usually regarded as an undesirable burden.

Are these judgements supportable by discoursal evidence? The claims are as much cultural as linguistic, and thus difficult to ‘prove’ with corpus evidence. But one can try. The phrase the stern task, for instance, occurs only twice in the half-billion-word Bank of English (to describe some daunting impending rugby matches-- Russia and Fiji have the stern task of facing England and Australia--and in the direct speech of a character in a novel set in wartime America: “… before we turn again to the stern task and the formidable years that lie before us…”). If there is no corpus evidence that living is a normal continuation of the phrase the stern task of, is there an equal lack of evidence of people completing the phrase living is with a stern task or similar pejorative or negative phrase? Here it is important to note that the Joycean sentence uses the phrase living, and not earning a living, securing a living, making a living or any other such narrowing formulation. It is living itself which the narrative calls a stern task to be achieved; and in doing so, I am arguing, words are co-selected unusually (what Hoey [2005] would call unprimed and perhaps unnatural). In turn, this strangeness of wording is, to use a phrase explained presently, implicature-generating.

How then is the phrase ‘living is …’ used in the corpus? Since the Bank of English is insensitive, in searches, to capitalisation or sentence-initial words, a large number of hits concerning the here-irrelevant compounds cost of living and standard of living are among the 438 instances of ‘living is…’ found in the corpus. Once these and other inapplicable embeddings are stripped out, the dominant rightward collocate of living is is easy – often echoing the lyrics of the song “Summertime”; or one finds an ironical reversal of the clearly positive normal formulation: Summertime and the living is definitely not easy. It’s hard work getting a tan. Or the song phrase is played upon: Summertime... and the living is sneezy in an article about hay-fever. Nowhere in the corpus is it represented that living is, as a norm, an unpleasant struggle.

Turning to the British National Corpus (100 million words) and making a random search for 50 typical instances of ‘….living is…’ (there are 78 instances in all, in this corpus), again most instances involve a subcategorisation that makes
the example inapplicable (earning a living, urban living, Country Living, to give the first three examples). Again when these are discarded, instances of the bare nominalisation living followed by is are few, but almost always with positive rightward collocates: easier, cheap (in the good sense: inexpensive), very important, and (clearly half-joking) a risky business!

There is thus in the Joyce sentence a concentrated series of prosodic or evaluative reversals of ‘normal expectations’, which coming from an accomplished writer we may tend to attribute either to ironic or deceptive intentions, as Louw (1993) suggests. Readers experience a double rapid reversal in their text-processing here: achieved predicts nothing like the stern task, and then the stern task predicts nothing like living. It may be the ‘switchback’ effect here, the frequency or density of consecutive contra-expectancy or non-congruent prosodies, that makes the full phrase so striking and conducive to implicature-generation by the reader.

The mention of implicature here is intended to suggest some continuities between the corpus linguistic idea of blatant exploitative breaches of discourse prosodic norms and Gricean indirect communication. As noted in Section 3 of Chapter One, Grice (1975) elegantly explained the calculated indirectness of much linguistic communication. He argued that when – as they often do – a speaker or writer says something that is not straightforwardly relevant, truthful, succinct or informative, then it is usually because it is indirectly and complexly relevant and informative (in ways that the addresser trusts that the addressee can figure out). In literary stories, as in our conversations, there is an enormous amount of hinting that the recipient needs to notice. Therefore, alongside the discourse prosody notion, Gricean pragmatics (or its descendant, the ‘relevance’ theory of Sperber and Wilson (1995)) can be useful in helping to articulate how it is that we find a wealth of irony and commentary in the narratorial remark about Lenehan, that No one knew how he achieved the stern task of living, but his name was vaguely associated with racing tissues.

The phrase ‘achieved the stern task of living’, we infer, is much at odds with Lenehan’s own way of looking at the world, and much more evocative of the language of figures of authority and respectability, or of those who submit to a regime of service and submission. But what first draws our attention to it is the sense that it is a deliberate departure in manner from a simpler and more direct report such as no-one knew how he made a living, or no-one knew what work he did, or no-one knew how he kept body and soul together. In Gricean pragmatic terms, we assume that the narrator knows this, and we infer that, by calculatedly using this markedly less straightforward phrase, the narrator wants us to consider various more complex but off-record ideas: that living a decent productive life is for many a serious business and hard work, though Lenehan may have little sympathy with such views; that the hard-working sober life is not actually much fun; that some in
positions of power and authority (police, priests, bosses) tend to curb any easygoing attitudes of those they control, by ‘stern’ emphasis on life as a ‘task’; that Lenehan by contrast would perhaps like to have as little as possible to do with hard work or any work (the second half of the sentence clearly implicates that he does not have a regular job on a sports newspaper); that the nearest this gentleman of leisure gets to work is the laying of bets, the circulation of gambling tips, and the cadging of food, drink, and money. But none of this is explicitly narrated in the sentence, these being things we can reasonably ‘vaguely associate’ (to use the narrator’s own phrase) with what has been actually said. The example is another reminder that almost the first thing a language approach to verbal narrative has to emphasise is that a significant part of the telling is often achieved by the rich web of what is not said, but is probably implicated, in the language used. Before leaving this master sentence let us note, again, the appositeness of further details in the total construction, in the frame of alienation and anonymity: the ‘agent’ who knows so little of Lenehan’s lifestyle is no particular individual or friend, but No one; and it is not the man himself who associates – or rather “is associated” – with racing tissues, but only his name.

One way of reconciling the semantic-prosodic and Gricean analyses sketched above is to conclude that textual phenomena that might be explained in terms of presuppositions in a semantic-pragmatic analysis will tend to be explained in terms of priming and prosody in a corpus linguistic analysis. Or it might be more accurate to say that the one analysis entails and builds on the other, so that both are needed, at distinct levels: the corpus linguistic awareness (however tacit) of the unusual or ‘anti-primed’ precedes and licences the more elaborate and discursive but speculative conjecturing of implicatures.

4. Keyword analysis of “Two Gallants”

Using any of the automatic concordancers available, we saw in Section 1.2 above that it is quite straightforward to draw up a list of the most noticeably frequent lexical items in “Two Gallants”. Such a list includes the following: street (26 occurrences), young (14), face (13), know (12), walked (12), turned, eyes (11), head, right, time (all 10), went, corner, night, girls (all 9), road (8), side (8), friends, hand, air, knew, man, woman (all 7). But in terms of unexpectedness or distinctiveness of a text’s lexical choices, a listing of raw frequencies does not go far enough. It tells us that street is just twice as frequent as face in the story, and that corner occurs fewer times than face. But perhaps it is entirely normal for face to occur often in short stories, or in short stories in the early 20th century, given fiction’s interest in individuals and individual consciousness, and western (if not human universal)
assumptions about the face and its expressions as a window onto a person's mind, feelings, and values. Perhaps it is corner that is the most atypically frequently used item among these three. To make such judgements, a reliable and much larger comparator or 'reference' corpus must be used, alongside the text of current interest; such a corpus can tell us whether face is generally quite frequent in short story text (and just how frequent it is, in the reference corpus), whether corner is comparatively infrequent, and so on. Identification of a text's disproportionately frequent words, relative to a large comparator corpus used as a source of norms of word frequency, is what the WordSmith Keywords programme enables. My aim became that of combining, for purposes of text analysis, Keywords methodology (see e.g. Scott 2005 and Scott and Tribble 2006) with attention to the sentences in which those repetitions occur, on the assumption that these might be key sentences in the text (cf Hoey 1991).

First, Wordsmith's Wordlist function can be used to produce a list, by frequency and by alphabetical order, of all the words occurring in a text; this instantaneously shows all those words that are repeated, and how often, but does not rank the repetitions on any basis of significance. Then the Keywords tool is used: the Wordlist results for “Two Gallants” are computationally matched with the Wordlist results generated from a much larger reference corpus I was able rapidly to build for the purpose, comprising approximately 500,000 words of early 20th century fiction (stories and novels by Katherine Mansfield, Virginia Woolf, E.M. Forster, Ellen Glasgow, Edna Ferber, etc.). Keywords uses one of two statistical measures (the chi square test or log-likelihood: I used the latter) to identify those words in the target text (here “Two Gallants”) which, relative to the comparator corpus taken as a basis of general norms of word-frequency, are disproportionately frequent or infrequent. The table below reproduces the Keywords results for the story, in very much the format that the WordSmith programme displays them. In order, the seven columns give the keyword ranking of words, the keyword itself (in capitals), its numerical frequency in the target text and then what percentage of the text this comprises, the frequency and textual proportion of the same item in the reference corpus, and finally the numerical expression of the item’s log-likelihood keyness in the target text.

What emerge as the positive keywords, in order of keyness, are Corley, Lenehan, street, he, his, Corley’s, walked, Mac, and so on. Notice that Keywords counts and makes calculations about Corley and Corley’s separately, as it would the separate forms street and streets: lemmatisation, if judged desirable, can only be done after a first automatic processing of the text. Just one negative (comparatively rare) keyword was found (nearly three times more frequent in the reference corpus): she.
Table 1. Keywords in “Two Gallants”

<table>
<thead>
<tr>
<th>N</th>
<th>WORD</th>
<th>Freq. In TG and %</th>
<th>Freq. in Ref Corp and %</th>
<th>KEYNESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CORLEY</td>
<td>46</td>
<td>0</td>
<td>437.0</td>
</tr>
<tr>
<td>2</td>
<td>LENEHAN</td>
<td>29</td>
<td>0</td>
<td>275.4</td>
</tr>
<tr>
<td>3</td>
<td>STREET</td>
<td>26</td>
<td>113</td>
<td>114.8</td>
</tr>
<tr>
<td>4</td>
<td>HE</td>
<td>163</td>
<td>7,454</td>
<td>103.5</td>
</tr>
<tr>
<td>5</td>
<td>HIS</td>
<td>112</td>
<td>4,725</td>
<td>81.6</td>
</tr>
<tr>
<td>6</td>
<td>CORLEY’S</td>
<td>6</td>
<td>0</td>
<td>56.9</td>
</tr>
<tr>
<td>7</td>
<td>WALKED</td>
<td>12</td>
<td>135</td>
<td>33.1</td>
</tr>
<tr>
<td>8</td>
<td>MAC</td>
<td>4</td>
<td>1</td>
<td>33.0</td>
</tr>
<tr>
<td>9</td>
<td>GINGER</td>
<td>5</td>
<td>6</td>
<td>32.4</td>
</tr>
<tr>
<td>10</td>
<td>TRAM</td>
<td>5</td>
<td>6</td>
<td>32.4</td>
</tr>
<tr>
<td>11</td>
<td>TOWARDS</td>
<td>9</td>
<td>70</td>
<td>30.6</td>
</tr>
<tr>
<td>12</td>
<td>STEPHEN’S</td>
<td>3</td>
<td>0</td>
<td>28.5</td>
</tr>
<tr>
<td>13</td>
<td>COMPANION’S</td>
<td>3</td>
<td>0</td>
<td>28.5</td>
</tr>
<tr>
<td>14</td>
<td>MERRION</td>
<td>3</td>
<td>0</td>
<td>28.5</td>
</tr>
<tr>
<td>15</td>
<td>PEAS</td>
<td>4</td>
<td>4</td>
<td>26.9</td>
</tr>
<tr>
<td>16</td>
<td>OF</td>
<td>131</td>
<td>9,328</td>
<td>25.2</td>
</tr>
<tr>
<td>17</td>
<td>SHE</td>
<td>21</td>
<td>6,575</td>
<td>31.1</td>
</tr>
</tbody>
</table>

It is worth noting that other easily-accessed programs besides WordSmith can speedily identify the comparatively disproportionately frequent (and infrequent) words in a text, one such being the Wmatrix program, hosted by Lancaster University (see Rayson 2003, 2007). And by way of departure from the standard tabulated listing of keywords, Wmatrix now offers a “Key visualisation” or “cloud”, in which an alphabetical listing of a text’s keywords alone is displayed, in quasi-textual format (i.e. putting as many keywords on each line or row as can fit), but with each word in an Arial type size representative of its keyness. Moving the mouse over a particular keyword displays its frequency and log-likelihood keyness.

The Keyword list of items from “Two Gallants” can be interpreted in various ways. For example, street might be picked out as the most key of the lexical items here (i.e. discounting proper names), or walked might attract attention, as the most key of the lexical verbs (in fact it is the only verb keyword identified when the log likelihood measure is applied). But there is more than a suspicion that the list is under-informative because the text is so short, and yet the list includes items like Mac and tram and peas, which are peripheral to the narrative.
4.1 Using keywords on “Two Gallants”: section by section

What emerges when “Two Gallants” is subdivided into the thirteen short segments (described at 2.2 above), and these thirteen segments are then analysed for their keywords (against the 20th century fiction reference corpus)? Here, exceptionally, the chi-square measure was used rather than the more selective log-likelihood one: when used on such small text samples as were involved here (300 words each) the log-likelihood computation identifies few if any keywords at all. With the minimum frequency set at 2 words, a manageable quantity of keywords emerge for the thirteen subsections (again, raising the threshold, to demand three instances within just 300 words would have excluded many of the items listed below).

1. streets, warm, grey, evening, air.
2. companion's, waterproof, slung, waves, expression, his, face
3. Lenehan, Corley, biscuit, noiselessly, takes, tongue, thinks, fine, street, night, told, she'd
4. Corley, earnestly, large, friend, gaze, passing, body, side, person
5. Corley, Lenehan, lothario, tram, she's, game, said, pull, damn, fro, right, eh, bit, used, know, running
6. notes, street, time, walked, each
7. Corley, Lenehan, leg, street, corner
8. Lenehan, Corley, belt, pace, approached, woman's, walked, towards, body
9. railings, ginger, shop, street, played
10. peas, Corley's, ginger, beer, mechanic, eaten, plate, girls spirit, he, girl, friends, less, bring
11. Corley, Mac, Lenehan, Westmoreland, Egan's, Street, George's, lamp, city, return, he, crowd, ten, seen
12. Corley, steps, result, cigarette, towards, running, stopped, wondered, walk
13. Corley, Lenehan, hallo, friend, towards

What these listings do not show is the degree of collocation among these keywords. Thus not only are streets, warm, grey, evening and air disproportionately prominent in the story’s opening paragraph, they also co-occur in close proximity. This is most noticeable and entirely deliberate in the first paragraph, which begins with The grey warm evening…warm air… and two sentences later refers to the warm grey evening. Sometimes the words collocate with themselves, as when sentence (1) ends by mentioning the streets and sentence (2) begins The streets, shuttered… This degree of keyword collocation is perhaps less remarkable in the opening paragraph, a palpably ‘wrought’ description. But it is found in many of the other sections too, where the few keywords of a section tend to co-occur closely, and are sometimes even phrase-mates, rather than being randomly dispersed. In this story at least, the
disproportionately recurrent words tend to recur together. There is a kind of concentration of their markedness in particular pockets of the text.

Could one deduce the plot, and tell the basic story of “Two Gallants”, if one were given only the thirteen listings of section keywords displayed above? Only in the vaguest and most random way, I suspect. As a highlighting tool, on these sizes of sample, Keyword analysis can tell us only a little concerning the ‘aboutness’ of the story, and less about its progression. These keywords are not outstandingly key, relative to their ambient text, and should not be interpreted as if they were; figuratively speaking, they are various low hills in an undulating landscape of many small hills, some of which are not listed against the thirteen sections because they are just a little less prominent than those selected. In the next chapter, by contrast, I suggest that focussing especially on those sentences containing the top or key keyword, the single most disproportionately frequent word in the text, may be more promising and revealing.
Chapter 3 showed some of the ways in which corpus tools focussed on lexical trends in a story can reveal something of a story’s themes and movement. But the persistent challenge is that of linking lexical repetition with narrative progression and structure. This chapter presents evidence which suggests that a story’s most key keyword, which is very often the proper name of a main character, may have an important story-structuring role. If this tendency is as widespread as these first analyses suggest, one implication is that writers are tacitly aware of the guiding function that repeated use of a character’s proper name can have. This chapter is the first of three which, using findings uncovered by corpus analytic methods, explore further the functions of repetitions and near-repetitions in literary stories. Such repetition patterns are important to narrative progression, even if narrativity is fostered by other means as well (as will be presented in Chapters 7 and 8).

1. A top keyword abridgement of “Two Gallants”

If we return to consideration of the keyword listing for the entirety of “Two Gallants” (as recorded in Table 1, near the close of Chapter 3), there is no avoiding the evidence that the top keyword is Corley, the name of the self-regarding exploiter of women who takes or borrows money from his woman friend at the instigation of his leech-like associate Lenehan. At first glance it goes against one’s intuitions to set much store by the fact that Corley and Lenehan are keywords in the story. One would expect the names of main protagonists to come high in a keyword list for any narrative, since keyword identification is based on comparison with other comparable texts where the ‘words’ Corley and Lenehan are unlikely even to be mentioned (they are, fleetingly, in Joyce’s Ulysses). Still, the fact remains that Corley is the most key keyword in “Two Gallants”, and rather than simply ignoring this proper name keyword (as I had in previous story analyses: e.g. Toolan 2004b) I decided to explore whether it might have some significance. Bearing in mind that the narration has no overt first-person narrator, so that Corley in principle could be denoted by pronoun or various definite descriptions – the policeman’s son, the stouter gallant, the stocky narrator, her companion – could those sentences in which the keyword Corley figures constitute some kind of textual frame or organisation?
What happens if just those sentences in which the 52 namings of Corley occur (46 instances of Corley and 6 of Corley’s) are extracted from the full text? Below is the text that emerges from such an automatic selection, 51 sentences (sentence 39 mentions Corley twice), numbered for easy reference, and with each next sentence beginning on a new line:

(1) “And where did you pick her up, Corley?” he asked.
(2) Corley ran his tongue swiftly along his upper lip.
(3) “I told her I was out of a job,” said Corley.
(4) Corley’s stride acknowledged the compliment.
(5) Corley was the son of an inspector of police and he had inherited his father’s frame and gut.
(6) As the two young men walked on through the crowd Corley occasionally turned to smile at some of the passing girls but Lenehan’s gaze was fixed on the large faint moon circled with a double halo.
(7) At length he said: “Well... tell me, Corley, I suppose you’ll be able to pull it off all right, eh?”
(8) Corley closed one eye expressively as an answer.
(9) “She’s all right,” said Corley. “I know the way to get around her, man. She’s a bit gone on me.”
(10) But Corley had not a subtle mind.
(11) “First I used to go with girls, you know,” said Corley, unbosoming: “girls off the South Circular.”
(12) “And damn the thing I ever got out of it,” said Corley.
(13) “Only off of one of them,” said Corley.
(14) “There was others at her before me,” said Corley philosophically.
(15) “You know you can’t kid me, Corley,” he said.
(16) “Honest to God!” said Corley. “Didn’t she tell me herself?”
(17) “Time enough,” said Corley. “She’ll be there all right. I always let her wait a bit.”
(18) “Ecod! Corley, you know how to take them,” he said.
(19) “I’m up to all their little tricks,” Corley confessed.
(20) Corley swung his head to and fro as if to toss aside an insistent insect, and his brows gathered.
(21) But Corley’s brow was soon smooth again.
(22) “There she is!” said Corley.
(23) “Let’s have a look at her, Corley,” he said.
(24) Corley glanced sideways at his friend and an unpleasant grin appeared on
his face.
(25) “O... A look at her?” said Corley, more amiably.
(26) Corley had already thrown one leg over the chains when Lenehan called
out: “And after? Where will we meet?”
(27) “Half ten,” answered Corley, bringing over his other leg.
(28) Corley did not answer.
(29) Corley returned a salute to the air.
(30) As he walked on slowly, timing his pace to theirs, he watched Corley’s
head which turned at every moment towards the young woman’s face like
a big ball revolving on a pivot.
(31) The problem of how he could pass the hours till he met Corley again trou-
bled him a little.
(32) When he had eaten all the peas he sipped his ginger beer and sat for some
time thinking of Corley’s adventure.
(33) In his imagination he beheld the pair of lovers walking along some dark
road; he heard Corley’s voice in deep energetic gallantries and saw again
the leer of the young woman’s mouth.
(34) His friends asked him had he seen Corley and what was the latest.
(35) He replied that he had spent the day with Corley.
(36) He set off briskly along the northern side of the Green hurrying for fear
Corley should return too soon.
(37) He leaned against the lamp-post and kept his gaze fixed on the part from
which he expected to see Corley and the young woman return.
(38) He wondered had Corley managed it successfully.
(39) But the memory of Corley’s slowly revolving head calmed him somewhat:
he was sure Corley would pull it off all right.
(40) All at once the idea struck him that perhaps Corley had seen her home by
another way and given him the slip.
(41) Would Corley do a thing like that?
(42) They were walking quickly, the young woman taking quick short steps,
while Corley kept beside her with his long stride.
(43) He knew Corley would fail; he knew it was no go.
(44) Corley remained standing at the edge of the path, a little distance from the front steps.
(45) Corley turned and went towards her.
(46) The door closed on her and Corley began to walk swiftly towards Stephen's Green.
(47) He called out: “Hallo, Corley!”
(48) Corley turned his head to see who had called him, and then continued walking as before.
(49) “Hallo, Corley!” he cried again.
(50) Still without answering, Corley swerved to the left and went up the side street.
(51) Corley halted at the first lamp and stared grimly before him.

This abridgement reduces the 3,900-word, 231-sentence original story to a version (not a summary) of just 51 sentences or 700 words, thus approximately 20% of the original text. The most remarkable thing about this text is that while it lacks much of the texture of Joyce’s story, and is mainly about Corley, nevertheless it does still tell a story, with nearly all sentence adjunctions coherent. In fact, adjacent sentences here are often so coherent that one is in danger of overlooking the fact that they are not adjacent in the original.

Even when transitions at first seem awkward, as for instance between (1) and (2), where (2) seems an inadequate reply to (1), the process of reading on, of reading these dispersed sentences as a continuous text, enables the reader to treat (3) as a partial answer to (1), and prepared for by (2). A brief glance at the original text reveals that in fact (2) does immediately follow (1), while ten sentences of direct speech intervene before (3). Similarly, it is a surprise to find that (7) and (8) are adjacent in the original.

An example of this ‘transformed cohesion’ is presented by sentences (11) to (13):

(11) “First I used to go with girls, you know,” said Corley, unbosoming; “girls off the South Circular.”
(12) “And damn the thing I ever got out of it,” said Corley.
(13) “Only off of one of them,” said Corley.

There is little that is incoherent here, or more elliptical, than in the original, despite the fact that in Joyce’s story several lines of speech from Corley and a comment from Lenehan intervene between (11) and (12). Nor is there much awkwardness
apparent between (20) and (21), where an irritation that causes Corley’s brows to
gather in the first is answered by his brow being ‘soon smooth again’ in (21) – even
if we do not in this version learn what it is that briefly caused Corley’s brows to
gather – despite there being five excised intervening sentences.

Remarkable, also, is the degree to which the text remains coherent despite the
large quantity of direct speech in the abridgement (a much higher proportion than
in the original story); that aspect of maintained coherence seems to have been
considerably facilitated by Joyce’s tendency to refer to Corley with his proper name
while pronominalising Lenehan. As a result, when Lenehan uses his companion’s
name in direct speech addressed to the latter, and the latter’s contribution is tagged
said Corley (regardless of which of the men speaks first), then coherent exchanges
can be preserved. This is exemplified by (22) and (23), between which three sen-
tences of description of the woman are given in the original story:

(22) “There she is!” said Corley.
(23) “Let’s have a look at her, Corley,” he said.

If Lenehan did not have such a pronounced tendency to name his companion
when speaking to him (interpretable as a form of sycophancy), then such coherent
sequences as (22)-(23) would not arise, as the following amended pairs suggest.
Consider if (22) were followed by either (23’), with Corley removed, or (23”),
where he is replaced by Lenehan.

(22) “There she is!” said Corley.
(23’) “Let’s have a look at her,” he said.
(22) “There she is!” said Corley.
(23”) “Let’s have a look at her,” Lenehan said.

Since neither sentence (23’) nor sentence (23”) includes the Corley keyword, these
versions of (23) would not be admitted into the abridged version, and (22) would
be immediately followed – much less coherently-- by (24):

(22) “There she is!” said Corley.
(24) Corley glanced sideways at his friend and an unpleasant grin appeared on
his face.

As things stand, the entire sequence from (22) to (27) is remarkably coherent, and
central to it is (23), both a coherent response to (22) and importantly preparatory
to (25): without (23), (25) would be somewhat ‘stranded’ in its informational rel-
evance, as it echoes (23). All this encourages the supposition that Corley is not the
top keyword without good text-organisational reason.
This is not to deny that some transitions in the abridgement are rather more awkward than those noted so far. The first such occurs between (3) and (4), where the latter uses the anaphoric noun phrase the compliment to refer to some words or gesture whose content is not here recorded. A few links are imperfect but certainly workable transitions, as between (19) and (20), or (13) and (14). In the latter, a reader’s inclination is to interpret the her of (14) as co-referential with the one of them, which in turn links back to the girls off the South Circular of (11); and all these anaphoric interpretations are correct. (14) and (15) may look an awkward pairing – but not for the first time we face the irony of finding that this speech exchange is exactly as sequenced in the original story, with only the following narratorial report of Lenehan’s scepticism intervening: This time Lenehan was inclined to disbelieve. He shook his head to and fro and smiled.

Thus the number of non-mutually-relevant pairs, in my estimation is surprisingly few: (3) to (4), (9) to (10), (16) to (17), and (27) to (28). A mere four out of fifty. Even in those four problematic transitions, we should not overstate the incoherence. For example, (10) certainly reads as a poorly-designed continuation from (9), but (11) rapidly restores narrative order, being a reasonable resumption of the direction taken by (9); (11) also confirms that (10), a coda to Lenehan’s veiled mockery of his companion, is the outlier sentence.

As suggested earlier, not only is there a high degree of local coherence between individual adjacent sentences in this abridgement, there is also global coherence: it would be hard to deny that a certain kind of story is told by this 51-sentence version. It is very much an in medias res version, with rather more focus on Corley than in Joyce’s original. But it still contains an abridged middle section in which, we know, Lenehan is passing the hours (31) until Corley’s return, eating a frugal meal and enviously imagining the lovers’ dalliance. Most striking of all are the final sentences, from (36) to (51), which still in coherent fashion encapsulate Lenehan’s anxious and suspenseful swings of mood, between confidence in Corley and distrust of him. These sentences read like the abrupt jump-cuts, leaving gaps in the continuity, that a modern film version might effect; for example, that between (44) and (45):

(44) Corley turned and went towards her.

(45) The door closed on her and Corley began to walk swiftly towards Stephen’s Green.

Equally noteworthy is the ‘coherence’ or wholeness of those segments that the 51-sentence abridgement causes to be set aside. Perhaps most striking is total absence of the episode at the very middle of the story (and for some critics, at the story’s core), where Corley and Lenehan are left subdued from encountering a weary harpist, playing a mournful melody. The episode falls between what are
sentences (21) and (22) of the abridgement, but you would not know this, so complete is the excision. The abridged text is the poorer for the episode’s absence, of course; nor do I mean to question whether the harp-playing scene is fully woven into the texture of Joyce’s story, just as fully as the opening paragraph of the story, which is similarly completely absent from the top keyword abridgement. But this simple keyword-based abridgement does suggest that even a Joycean story reveals an underlying segmentation or organisation, a mediate structuring below the level of the whole story and above that of the individual sentence.

2. Applying the top keyword procedure in story abridgements

2.1 Updike’s “A & P”

Is the relatively coherent and narrative-like abridgement of “Two Gallants”, around its top keyword Corley, anything more than a local curiosity? To address this question, I applied the same procedure to a number of other stories, with broadly similar results, and I discuss some of these in this section. The first replication was attempted with a very short story, John Updike’s “A & P”. Against a 400,000 word twentieth-century fiction corpus, what emerged as the story’s keywords were the following:

<table>
<thead>
<tr>
<th>N</th>
<th>WORD</th>
<th>Freq. in A&amp;P and %</th>
<th>Freq. in Ref. Corp and %</th>
<th>KEYNESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LENGEL</td>
<td>10 0.35</td>
<td>0</td>
<td>94.7</td>
</tr>
<tr>
<td>2</td>
<td>AROUND</td>
<td>13 0.46</td>
<td>80 0.02</td>
<td>49.3</td>
</tr>
<tr>
<td>3</td>
<td>QUEENIE</td>
<td>5 0.18</td>
<td>0</td>
<td>47.3</td>
</tr>
<tr>
<td>4</td>
<td>SNACKS</td>
<td>5 0.18</td>
<td>0</td>
<td>47.3</td>
</tr>
<tr>
<td>5</td>
<td>SLOT</td>
<td>5 0.18</td>
<td>1</td>
<td>42.0</td>
</tr>
<tr>
<td>6</td>
<td>GIRLS</td>
<td>14 0.49</td>
<td>141 0.04</td>
<td>41.1</td>
</tr>
<tr>
<td>7</td>
<td>P</td>
<td>6 0.21</td>
<td>6</td>
<td>40.3</td>
</tr>
<tr>
<td>8</td>
<td>SAMMY</td>
<td>4 0.14</td>
<td>0</td>
<td>37.9</td>
</tr>
<tr>
<td>9</td>
<td>STOKESIE</td>
<td>4 0.14</td>
<td>0</td>
<td>37.9</td>
</tr>
<tr>
<td>10</td>
<td>HERRING</td>
<td>4 0.14</td>
<td>0</td>
<td>37.9</td>
</tr>
<tr>
<td>11</td>
<td>AISLE</td>
<td>5 0.18</td>
<td>7</td>
<td>31.2</td>
</tr>
<tr>
<td>12</td>
<td>KIND</td>
<td>11 0.39</td>
<td>133 0.04</td>
<td>28.8</td>
</tr>
<tr>
<td>13</td>
<td>SHOULDERS</td>
<td>7 0.25</td>
<td>36 0.01</td>
<td>28.7</td>
</tr>
<tr>
<td>14</td>
<td>PLAID</td>
<td>3 0.11</td>
<td>0</td>
<td>28.4</td>
</tr>
<tr>
<td>15</td>
<td>POLICY</td>
<td>3 0.11</td>
<td>0</td>
<td>28.4</td>
</tr>
</tbody>
</table>
The topmost keyword is *Lengel*, used ten times (less frequent but much more key than *around* or *girls*). To those ten I have added an eleventh instance where the contraction *Lengel’s* occurs, not caught by the form-sensitive original keyword search. With the aim of not contaminating the reader’s assessment of the putative coherence of the story abridgement, I defer any story-describing commentary until after presenting, now, the eleven-sentence abridgement based on that top keyword. The readers are invited to make their own assessment of whether this is coherent text or mere cutting and pasting of fragments:

(1) Lengel comes in from haggling with a truck full of cabbages on the lot and is about to scuttle into that door marked MANAGER behind which he hides all day when the girls touch his eye.

(2) Lengel’s pretty dreary, teaches Sunday school and the rest, but he doesn’t miss that much.

(3) “That’s all right,” Lengel said. “But this isn’t the beach.”

(4) “That makes no difference,” Lengel tells her, and I could see from the way his eyes went that he hadn’t noticed she was wearing a two-piece before.

(5) I could feel in the silence everybody getting nervous, most of all Lengel, who asks me, “Sammy, have you rung up this purchase?”

(6) The girls, and who’d blame them, are in a hurry to get out, so I say “I quit” to Lengel quick enough for them to hear, hoping they’ll stop and watch me, their unsuspected hero.

(7) They keep right on going, into the electric eye; the door flies open and they flicker across the lot to their car, Queenie and Plaid and Big Tall Goony-Goony (not that as raw material she was so bad), leaving me with Lengel and a kink in his eyebrow.

(8) “I don’t think you know what you’re saying,” Lengel said.

(9) Lengel sighs and begins to look very patient and old and gray.

(10) “You’ll feel this for the rest of your life,” Lengel says, and I know that’s true, too, but remembering how he made that pretty girl blush makes me so scrunchy inside I punch the No Sale tab and the machine whirs “pee-pul” and the drawer splats out.

(11) Looking back in the big windows, over the bags of peat moss and aluminum lawn furniture stacked on the pavement, I could see Lengel in my place in the slot, checking the sheep through.

Does this eleven-sentence text make any sense as an abridged story? I submit that, somewhat surprisingly, it does.
Chapter 4. Top keyword sentences as story waymarking

Lengel is the surname of a supermarket manager who admonishes three teenage girls who have come into the store wearing only their swimsuits; Sammy, the impressionable teenager who narrates the story, is sufficiently upset by what he regards as Lengel’s heavy-handedness that he abruptly quits his job on the check-out counter. Sammy is particularly taken with one of the girls, for whom he privately invents the name “Queenie”. Again, then, a story of 2,900 words and 123 sentences has been automatically abridged, this time to the eleven sentences in which the name Lengel appears, a ‘text’ of 297 words, thus a reduction to approximately 10% of the original. This can be compared with the 51 sentence, 700-word abridgement of “Two Gallants”, a reduction to 20% of the original. Since the latter worked rather well, it may be that the Lengel-based abridgement of “A&P” is ‘too severe’ and that, for example, all the sentences containing the next most frequent keyword should be added to make a more coherent abridgement – we shall see, below.

It is finally for the individual reader to decide to what extent the eleven-sentence abridgement is a coherent text. Not wishing to presume more significance than was warranted for this tendency of the top keyword sentences to form a semi-coherent abridgement, I presented the eleven-sentence text to twelve colleagues and postgraduate students, to obtain some independent opinions. I received twelve sets of replies to the four questions I appended to the text.

My first question ran as follows:
1. Do you recognise or vaguely recognise this text?
Yes/No.

and the reply, from all twelve respondents, was No. The second question was:
2. Does this text strike you as a coherent or semi-coherent narrative?
Yes/ Semi-coherent/ No.

To this, only one respondent said “No; semi-coherent on second reading (filling in some gaps)”; six judged it semi-coherent, one adding ‘maybe a film script made more coherent by the action?’ another astutely adding “Bits of it seem to have been deleted.” Three said it was semi-coherent moving to coherent on a second reading, one commenting “A significant interruption is the sense of something missing between sentences 3 and 4. To make this coherent, I supplied the idea that the narrator could hear Lengel but not the girls.” Two respondents simply answered “Yes [coherent]”. In summary, two respondents rated the text fully coherent on a first reading, and after a second or third reading all were inclined to rate the text as at least semi-coherent. Some respondents referred in the course of their comments to ‘filling in some gaps’ – apparently something they routinely did.

My third question asked, in a deliberately vague way, “Do you have any sense of what the problem with the girls is?” Four of the twelve replies were wrong or wayward (“Yes – sexual harassment from Lengel” one offered), but the other eight
accurately sensed that inappropriately revealing clothing was the issue, with one respondent being remarkably to the point: “They’ve gone into the shop where Lengel and Sammy work wearing ‘inappropriate’ clothing – swimwear.”

My fourth and final question was intended to probe quite a dependent and more weakly-cued inference. I asked respondents “What does ‘the sheep’ in the final sentence refer to?” Even on this more difficult task, which stumped some readers, five of the twelve respondents hazarded that it referred to customers or (respondents’ own scare quotes) ‘anonymous people,’ ‘the masses.’

From these answers I deduce that the abridged text is, for almost all proficient readers, an at least semi-coherent narrative. Notice that this is a higher hurdle than simply judging it a semi-coherent text: readers have implicitly looked for a ‘before and after’ in the text, and most are of the view that it is apparent. Answers to follow-up Questions 3 and 4 were designed to put a bit of pressure – quite a bit of pressure one might argue – on respondents’ predicted answer to Question 2. That is, I wanted to see whether, for example, an informant who said the text was a fully coherent narrative nevertheless had unsatisfactory answers to Questions 3 and 4. But here we must be careful as to what we deem ‘unsatisfactory.’ The ‘right’ answer to Question 4, about who the ‘sheep’ were, is, by the lights of the original story, the store’s typical middle-aged (docile, boring) customers. Five out of twelve impressively gave this answer. But is there a danger here of unfairly judging the abridged text and its coherence by the standards and potentials of the original text, a quite different one? The answer must be that there is such a danger – but this is also a necessity, for the particular task and interest here, namely whether an abridgement such as the one here devised continues satisfactorily to represent a significant part of the bare plot and core progression of the unabridged text. It is interesting to notice, also, that a few of these informants (educated, articulate, and used to completing questionnaires), when asked what ‘the sheep’ referred to, were inclined not to stray too far from the literal and conventional meaning in their answer: ‘some kind of sheep, perhaps toy or model?’ one suggested. Sentence (11), with its late placement and therefore non-prominent mention of ‘the sheep’ that Lengel is checking through, does not openly signal its figurative and negatively-evaluating contextual meaning. The sentence also involves metonymy (someone operating a supermarket till checks through the items of the customers, not the customers themselves). Thus there is a density of creative language use here, and this makes it less surprising that most informants did not fully understand these hinted meanings. An unsatisfactory answer to Question 4, from an informant who has judged the narrative semi- or fully coherent, would be “I don’t know”. But even this might be rescued if we assume that such an informant does not mean to say “I have no idea, no understanding, at all” but only “I don’t know what particular thing is or are being referred to by sheep, although obviously some kind of things
are being referred to”. One little irony here: because of the unchanged plural form of *sheep*, the final sentence is actually ambiguous between ‘one sheep’ and ‘several sheep’ readings, and this ambiguity was maintained in Question 4; but all informants gave a plural interpretation.

Answers to Question 4 that show full or partial comprehension are perhaps a secondary confirmation of the text’s semi-coherence if not its ‘substantial’ coherence. But the answers to Question 3, which probes for informants’ awareness of the ‘problem’ or crisis which provides the dynamic of the story, are arguably criterial confirmation of prior assessments of the story as at least semi-coherent. And here, just over half of the informants hazarded that the girls’ inappropriate dress was the issue, with only four of twelve seemingly ‘lost’. Still it is hard to deny that the proportion of correct answers to Question 3 is somewhat lower than one would like, by way of confirmation that a narrative abridgement is a clear and adequately full one. And this in turn forces one to recognise that the abridgement simply doesn’t indicate, in a fully adequate way, that three girls have come into a town-centre store in their beachwear. Reflecting on “A & P” more generally, one might argue it has two phases: first, the entry of the young and semi-naked girls and the varying reactions of the narrator, customers, and manager; then, the narrator’s romantic-melodramatic ‘protest’. A summary shaped around all mentions of the manager by name – the manager not being the central figure of either episode – can catch the gist of the second episode and some at least of the first, and that is still of considerable interest. But it isn’t fully successful, for reasons that need to be clarified: they may have to do with the perspectival shift when Sammy moves from being chiefly an observer of events to being – with his “I quit” – a prominent actor.

2.2 Carver’s “Boxes”

In some ways the top keyword abridgements that emerged for the two stories analysed next were more successful; those stories, both by Raymond Carver (Carver 1983 and 1988), were “Boxes” (top keyword: *Jill*) and “Cathedral” (top keyword: *blind*). “Boxes” is composed in five distinct sections, and due to space constraints I reproduce here just the abridgement of the opening section:

“Come on and eat with me one last time,” she says. “You and Jill.”
So finally I turn to Jill and say, “Let’s go to my mother’s for a good-bye meal.”
Jill is at the table with a Sears catalogue in front of her, trying to find us some curtains.
Jill always says what’s on her mind.
Jill had to borrow money to bury him, and then – can you beat it? – she was presented with a bill for the bridge repair.
“Hey, Jill, don’t do any favors. Do you want to come with me or not?”
Jill tried to make the best of it. Jill said it reminded her of the situation with her first husband’s mother. “She was a clinger,” Jill said. It’s fair to say that my mother sees Jill as an intruder. As far as she’s concerned, Jill is just another girl in a series of girls who have appeared in my life since my wife left me. Jill and my mother act friendly enough when they find themselves together. But Jill dreads the time she has to spend in my mother’s company. After a while, ten minutes or so, Jill comes out of the bathroom. Jill doesn’t look at me. Jill says, “It’s okay.”

This is no longer Carver exactly, but neither is it total gibberish. Perhaps I am too predisposed to find this text – give or take an inference here and there – largely coherent and poignant in its cut-down-ness, simply because I already know the Carver story well. Several sentences clearly do not sit well, cohesively, relative to what has gone before, or after, or both, the first really troublesome sentence being the fifth one, about burying ‘him’. But it is still striking that such entirely aberrant sequences are so infrequent. Though these things are hard to measure, my impression is that summarising via the top keyword is more challenging in “Boxes” than in “Two Gallants”, perhaps particularly because there is more direct speech in “Boxes”, and the name Jill is used quite often in that direct speech. In principle such direct speech could be uttered by a speaker removed in space and time (e.g., on a telephone message): such deictic shifts add to the processing burden on the reader.

Now consider the abridgement of the fifth and final section of the story, a spare three sentences, since the top keyword Jill is infrequent here. The 520 words of this section of the story is reduced to just 32 words:

Jill is at the table with the catalogue, listening.
Jill turns the pages of her catalogue, and then she stops turning them.
“What is it you see out there, honey?” Jill says. “Tell me.”

This does not and cannot retain the implication-evoking complexity, the richness of situation, of the original; but it does capture and highlight some of the shaping choreography of that final scene. In Emmott’s terms (Emmott 1997) the contextual frame of these three textually-scattered sentences remains constant. The abridgement foregrounds the fact that these three sententially-represented phases or moves are connected to each other (not just by the shared mention of Jill, but also the catalogue, the sense that Jill is still at the table in the third and final sentence, that her stopping turning the pages is linked to her final enquiry and request in ways not textualised further in the full text – e.g., that she stops looking at the
catalogue in order to look at the narrator, that she sees him looking “out there”, and that this prompts her to ask the question that we also might reasonably ask.

The five-section abridgement of “Boxes” is a purely mechanical extraction of those sentences of the story in which the top keyword, Jill, appears. Over 5,000 words, nearly 500 sentences, are thereby reduced to just 730 words in 55 sentences. Absent from it is the original story’s resolution between son and mother, where he sees how he can and should start calling her dear, discussed below. But what is not lost is a charting of the tense, partially-unresolved relations between the three people; and, as indicated, it ends almost as powerfully and openly as the original, with Jill at the table thumbing through her catalogue, the narrator looking out at or for something undeclared, that we, like Jill, are not quite sure about but think we know.

Consider, similarly, the following abridgement sequence, which seems to me remarkable not only for being largely coherent, but also for what it manages to convey – namely, something of that poignant, inarticulate, fumbling interpersonal connection between mother, son, and daughter-in-law (six relationships) that Carver captured in the original story:

Then she says, “I’ll miss you, too, Jill.”
Jill sips from her coffee and nods.
“I wish you could have been happier here,” Jill says.
“Jill,” I say.
She puts her cup on the floor next to her chair and waits for Jill to tell her she isn’t asking for too much. But Jill doesn’t say anything, and in a minute my mother begins to outline her plans to be happy. After a time Jill lowers her eyes to her cup and has some more coffee.

Inevitably, important kinds of material are lost. And interestingly or ironically enough, one of the things that is lost is the repetitive use of the word “dear” (always within quotation marks) by the narrating son, who is on the telephone to his mother, towards the story’s close – a verbalised long-distance counterpart of a reassuring touch. Dear is used quite frequently in the final section, and yet it does not emerge as a keyword for the story as a whole. Another word that is completely lost is boxes, which occurs nowhere in the fifty-five sentences that include the name Jill: but then boxes are not very prominent in the original story, beyond the title. Whether ‘boxes’ is also lost as a story theme in the abridgement is a harder question to answer.

There are eight dears, 0.15% of text. But so, too, are there 8 honeys and 8 coffees. In the comparator corpus, instances of dear amount to just 0.07% of the text (331 tokens in nearly half a million), so dear is twice as frequent in the Carver story as in what I have used as my literary fictional norm, and twice as frequent is just not frequent enough to stand out as a keyword in the story. But again, one can see that
things could be looked at differently; in particular one could subdivide the story into plausible sections, and only then compute its keywords. If one did that, then dear would undoubtedly emerge as locally key, in those few sentences where the son recalls his father’s term of affection and then uses it to his mother. It seems quite reasonable to say that dear is prominent and important to the texture of just that important passage in the story, but not contributory to the major spine, the plotline of the story, in the way that jill, arguably more than any other word, is.

2.3 Carver’s “Cathedral”

As a final short story example in this section I will briefly discuss the top keyword abridgement of Carver’s story “Cathedral”. Overall, abridgement of “Cathedral”, around the eighty sentences using the key keyword blind, seems the most successful of those attempted so far. The abridgement is of particular interest since here for the first time the top keyword is not a proper name as in the previous analysed stories, but an adjective. What is immediately apparent, however, upon studying the story is that blind occurs in the vast majority of instances in the phrase the blind man, the formulation that the (unnamed) first-person narrator almost invariably uses to refer to the central character (whose given name, Robert, is rarely used). So the top keyword blind functions, in practice, within a name-like phrase. Here is how the story abridgment opens:

This blind man, an old friend of my wife's, he was on his way to spend the night. But she and the blind man had kept in touch. And his being blind bothered me. In the movies, the blind moved slowly and never laughed. A blind man in my house was not something I looked forward to.

She'd seen something in the paper: help wanted – Reading to Blind Man, and a telephone number. She'd worked with this blind man all summer. They'd become good friends, my wife and the blind man. On her last day in the office, the blind man asked if he could touch her face.

Again it is surprising how neatly this version knits together. Consider, for example the two sentences below, which are several lines apart in the original story, so that there seems no guarantee that they will cohere. Uncannily, they mesh remarkably well here, the rheme or comment of the first (the blind man's wife) being co-referential with the theme or topic of the second (Beulah).

She’d told me a little about the blind man’s wife.
Beulah had gone to work for the blind man the summer after my wife had stopped working for him.
Is this mere felicitous coincidence, or is there some sensitivity to narrative pattern-
ing by means of high frequency name that is ordinarily ‘below’ the reader’s normal
awareness (but emphatically not, it seems, that of great writers)?

Another powerful textual progression that the abridgement creates is the fol-
lowing:

“Do I look distinguished, my dear?” the blind man said to my wife.
My wife finally took her eyes off the blind man and looked at me.

A later transition adds a pleasing twist to what was already an amusing moment in
the original. This occurs when the woman falls asleep, while sitting on the sofa
between the two men, in such a way that (her husband notices) her night robe
“had slipped away from her legs, exposing a juicy thigh”. Prompted by decorum or
proprietary, the husband reports “I reached to draw her robe back over her, and
it was then that I glanced at the blind man. What the hell! I flipped the robe open
again”. In the abridgement this generates the following transition:

I reached to draw her robe back over her, and it was then that I glanced at the
blind man.
“Bub, it’s all right,” the blind man said.

Something that this exercise powerfully confirms is the way the word _blind_ drops
away from prominence, as the story progresses. This can also be shown using the
dispersal graph in WordSmith Tools to plot the distribution of a word through the
course of a text.

The eighty-sentence abridgement of “Cathedral” is arguably a little more het-
erogeneous than the counterpart ones built around the proper names _Jill, Corley_,
etc. By virtue of the diverse uses of the form _blind_, as adjective or noun, with spe-
cific or generic reference, some sentences are retained in the abridgement which
one might otherwise want to exclude, among them this late piece of stupidity (in
the final section) from the husband:

A wink is the same as a nod to a blind man.

Even here, however, one pauses to wonder. In the abridged version this last gasp of
prejudice is immediately followed by the report that the blind man sits very still,
_listening_ to the husband, before clearing his throat and getting down to the business
of drawing the cathedral (the fact that Robert and the narrator hold hands to per-
form this task is sadly lost in the abridgement). Could it be, uncannily, that Robert
somehow hears or senses the husband’s reported but unspoken disparagement?

In this section I have suggested that the top keyword abridgement may be re-
vealing of the linear and developing structure of a range of short fiction, showing
how it applies, to a degree, first to “Two Gallants” and then to several other stories.
Repeatedly, informants have judged the derived texts to be at least “semi-coherent”, and have been able to answer questions about under-described events and characters by means of intelligent inferential elaboration of textual clues. In the next section I briefly address some procedural questions about the top keyword method, before commenting in Section 4 on what the abridgements might mean.

3. Further procedural questions about the top keyword method

3.1 Distinguishing the top keyword from other frequent keywords.

Does keyword-abridgement work best with the top keyword? Is an abridgement using only sentences using the second or third most key keyword in “Two Gallants” or “A & P”, or an abridgement that combines all uses of the first and second most key keywords, substantially inferior in coherence and narrativity to the top-keyword-only version? My studies so far suggest that all these questions must be answered in the affirmative. In the case of “A & P”, an abridgement built around the third most key keyword (Queenie), although it does not tell much of a story, retains some coherence even if it leaves much to be guessed or inferred, as this sample shows:

> Around they come, Queenie still leading the way, and holding a little gray jar in her hand.
> Queenie puts down the jar and I take it into my fingers icy cold.
> Queenie blushes, though maybe it’s just a brush of sunburn I was noticing for the first time, now that she was so close.
> Queenie’s blush is no sunburn now, and the plump one in plaid, that I liked better from the back – a really sweet can – pipes up, “We weren’t doing any shopping. We just came in for the one thing.”

But an abridgement based solely on the second top keyword (around) is frequently incoherent, especially in the latter half. Similarly, an abridgement based on the top two keywords (Lengel and around) is less coherent and satisfactory than the eleven-sentence abridgement based on Lengel alone, presented earlier. Sequences such as the following arise:

> Lengel’s pretty dreary, teaches Sunday school and the rest, but he doesn’t miss that much.
> Her father and the other men were standing around in ice-cream coats and bow ties and the women were in sandals picking up herring snacks on toothpicks off a big plate and they were all holding drinks the color of water with olives and sprigs of mint in them.
> “That’s all right,” Lengel said. “But this isn’t the beach.”
Around is overwhelmingly adverbial rather than prepositional in this text, and appears in a range of uses: there is no element of logical continuity or recurrence or chaining in its use. By contrast these are some of the qualities more or less guaranteed of a proper name keyword in a narrative text: that, typically, each recurrence involves a ‘next important thing’ in the telling which involves the entity so named, with an expectation that time has passed (and perhaps other changes have occurred) since the previous mention.

As for “Two Gallants”, an abridgement around the story’s second keyword, Lenehan, despite the word occurring in twenty-nine sentences, turns out to be disastrous on its own, with incoherent sequences such as the following not untypical:

Lenehan said no more.
Lenehan grew lively.
"Damn it!" said Lenehan boldly, "I don’t want an introduction. All I want is to have a look at her. I’m not going to eat her."
"Right!" said Lenehan.
Corley had already thrown one leg over the chains when Lenehan called out: "And after? Where will we meet?"
"Work it all right now," said Lenehan in farewell.

If the Lenehan sentences are combined with the Corley ones, we get a reasonably coherent 78-sentence abridgement of the story. For some sequences, this works well, particularly where it reports Lenehan-Corley dialogue fairly extensively; it is perhaps most effective as a report of the story’s final episode. But at other points there is quite a degree of incoherence, as in the following sequence:

(55) The problem of how he could pass the hours till he met Corley again troubled him a little.

(56) When he had eaten all the peas he sipped his ginger beer and sat for some time thinking of Corley’s adventure.

(57) In his imagination he beheld the pair of lovers walking along some dark road; he heard Corley’s voice in deep energetic gallantries and saw again the leer of the young woman’s mouth.

(58) His friends asked him had he seen Corley and what was the latest.

(59) He replied that he had spent the day with Corley.

(60) At this Lenehan said that he had been with Mac the night before in Egan’s.

(61) Lenehan did not know: he said that Holohan had stood them drinks in Egan’s.
On balance, despite the addition of twenty-seven extra *Lenehan*-mentioning sentences, this longer abridgement seems less coherent than the *Corley* text. Why should this be? It may stem from the fact that Lenehan is the centre-of-consciousness for a good deal (but not all) of the narration, when he is typically denoted by a pronominal referent rather than his name, so that naming him by proper name is not a consistent practice (as it is, to a much greater degree, in the case of Corley). Added to this, Lenehan is the respondent in this story, while Corley is frequently the initiator.

3.2 On the importance, in the top keyword, of proper name status.

An early question to arise concerned whether the keyword involved has to be a proper noun or name, as it was in the first stories so analysed. Does the procedure yield a semi-coherent text when applied to a story which lacks as top keyword a proper name? These questions require fuller exploration than is possible here, but an interesting first adjustment is prompted, as we have seen, by the case of “Cathedral”. In that story,-blind (and not a name) is by a clear margin the topmost keyword, and the generated abridgement is remarkably coherent. In a sense this suggests that the top keyword need not of necessity be a proper name for a coherent abridgement to emerge. On the other hand, in the vast majority of uses in that story, as noted earlier, blind is part of the phrase the blind man (sometimes this blind man) and denotes a main character who is referred to thus, rather than with his personal name (Robert), by the threatened and stereotyping narrator. Thus the top keyword blind is overwhelmingly used in the narration with a character-denoting function. In the story it functions much like a proper name: it is the main way of naming an important, frequently focalised, character. It may also be worth reporting that a character name, the focalised or the second most important but intermittently focalised character, does often seem to emerge as top keyword: in “Bliss” by a large margin it is Bertha, in “The Love of a Good Woman” it is Enid, in Munro’s “Royal Beatings” it is Flo, and so on.

A greater complication is presented by a story such as Joyce’s “Eveline”, where the top keyword is a pronominal form (her), and therefore has potential for use to refer to a variety of different female participants. As much could be said, in principle, about blind, by contrast with the story norms for personal names (story names like Lengel and Corley tend to be uniquely-refering, although they do not have to be). But as it happens again the vast majority of uses of her refer to Eveline, and there are such a considerable number of them that an abridgment comprising all the sentences in which her occurs is coherent but somewhat unrevealing. It may be that where the top keyword (whatever its nature, but pronouns are the most likely
candidates) is so frequent, occurring in as many as half a text’s sentences, a limiting point in the usefulness of keyword abridgements has been passed.

4. Some interim conclusions

What can we say about Lengel in relation to “A&P”, or about Corley in relation to “Two Gallants”? Simply because each has emerged as the top keyword in the story where they feature, we can argue that each is the most foregrounded word in their respective texts. Here perhaps a word or two is in order concerning what is meant by foregrounded language and foregrounding generally.

The transfer of the terms foreground and background, and of figure and ground, from references to depicting with perspective to figurative uses in many everyday situations, is commonplace. But foregrounding and foregrounding theory are also cornerstones of literary stylistics, adopted first from Russian and Prague School formalists, especially Mukařovsky ([1938]1964) and elaborated in studies that are increasingly attentive to psychological or cognitive factors (see van Peer 1986; Douthwaite 2000; van Peer, Zyngier and Hakemulder 2007; and a recent special issue of the journal *Language and Literature*: van Peer 2007a and 2007b)

Foregrounding in the literary text, by means of some form of noticeable excess, markedness, prominence, contrast, or departure from one or more of the text’s multiple norms (the text’s typographical, or rhythmic, or dialectal, or syntactic, or pragmatic, etc. norms) is assumed to be a fundamental sense-making aesthetic resource that bears directly on the text’s reception, interpretation and evaluation as literature.

If it is true, then, that the top keyword in stories is the most foregrounded lexical item, this is no minor attribute. But in suggesting that top keyword abridgements may constitute foregrounded story material, I am not claiming that these are necessarily the core or some of the most important sentences of a narrative text. It might be more useful to draw an analogy with an archaeologist picking over the relics of some pioneer’s log cabin: when all other parts have biodegraded, all that may remain might be the metal hardware of hinges and handles, where once the windows and doors were situated. The top keyword sentences may be similarly perdurable elements in the story structure.

Top-keyword abridgements are mere shadows of the original stories. But they are neither plotless nor incoherent. In evaluating one, the analyst may be too influenced by their knowledge of the original, therefore too inclined to find it – give or take an inference here and there – largely coherent and expressive in its reduced condition. It is therefore important to present the abridgement to independent reader-judges who are unfamiliar with the original. Still, *entirely* aberrant se-
quences are infrequent. Among those stories I have analysed in this way thus far, those involving extensive direct speech tend to give rise to a less smoothly coherent abridgement. This may be because all forms of reported discourse can involve change of speaker or time or place, so that any abridgement of a text that omits these kinds of deictic guidance will present interpretive difficulties, perhaps insurmountable ones. More generally, the more time-shifts or space-shifts that are involved in a story (particularly analeptic shifts) the less secure its top keyword abridgement is likely to be.

Abridgements lack the resonating texture, the breadth, of their authorial originals. But a top keyword abridgement of a Joyce story has something eloquently Joycean about it, and one of a Carver story something Carveresque, in ways that a summary would not. This is not entirely surprising, in view of the fact that the abridgement comprises only sentences composed by Joyce, Carver, or whomever. Abridgements (largely) make sense – not just banal sense of the type in which a chronological sequence is accurately reported, but in addition the kind of textured story sense to which as readers we feel able and willing to contribute inferences, hazard implicatures, and begin to feel emotional alignments with one or another character. This is quite unlike what happens in the creating or reader processing of a standard summary of a narrative. The abridged 51-sentence version of “Two Gallants” can be read as verbal art; the Jakobsonian poetic function, or literariness, survives the manipulation. Reading an abridgement is (or may be) experiential or reader-involving in a way that an orthodox summary is not.

In the reading of a standard story- or novel-summary, little or no experiential involvement in the text, as the projection of a situation, is felt by the reader. In fact summaries are worded and designed to communicate without generating such involvement. They are also usually quite short, and one for a short story might easily fit on one page of text. But imagine such a short summary (of, say, twenty sentences) presented in such a way that it was distributed across ten pages (or screen images), with just two sentences per page or screen, so that the reader could not too easily take in the whole summary at once. Such a contrivance would confirm, I believe, that no-one, at any point in a summary, feels empathy for one character, antipathy for another, and thinks “Oh, what will happen next?” A good summary of “Two Gallants” or “Boxes” is not to any degree a work of verbal art, nor is it intended to be; it will have taken all reasonable steps to banish ambiguity, and is unlikely to prompt inference-making or the calculation of implicatures. By contrast the reading of an abridgement can be experiential, involving, and quasi-artistic, making us alert to potential ambiguities and opacities. This inauthentic, mechanically-generated, top-keyword abridgement has many of the features and resources, and therefore some of the effects, of the full-length authentic original.
The abridgements discussed above seem to suggest that to some degree writers ‘orient’ or organise an important part of the skeleton of the story around all and only those sentences containing the most exceptionally foregrounded word. Is this the more apparent if, for example, we examine a version of “Two Gallants” from which the fifty-one ‘Corley’ sentences have been removed? Here is the closing of the story, with the final eight ‘Corley’ sentences removed. This amended closing passage, comprising twenty-four sentences, is three times the length of the equivalent section of the abridgement, and conveys much more of the original’s texture. Nevertheless it leaves the reader much more confused or undirected as to precisely who is doing what to whom at the story’s close. The basic narrative trajectory is obscured here, whereas it is exceptionally vivid in the mechanically-derived eight-sentence abridgement (sentences 44 to 51 of the abridgement supplied above in 4.2).

They turned down Baggot Street and he followed them at once, taking the other footpath. When they stopped he stopped too. They talked for a few moments and then the young woman went down the steps into the area of a house. Some minutes passed. Then the hall-door was opened slowly and cautiously. A woman came running down the front steps and coughed. His broad figure hid hers from view for a few seconds and then she reappeared running up the steps.

Lenehan hurried on in the same direction. Some drops of light rain fell. He took them as a warning and, glancing back towards the house which the young woman had entered to see that he was not observed, he ran eagerly across the road. Anxiety and his swift run made him pant.

Lenehan ran after him, settling the waterproof on his shoulders with one hand.

He came level with his friend and looked keenly in his face. He could see nothing there.

– Well? he said. Did it come off?

They had reached the corner of Ely Place. His features were composed in stern calm. Lenehan kept up with his friend, breathing uneasily. He was baffled and a note of menace pierced through his voice.

– Can’t you tell us? he said. Did you try her?

Then with a grave gesture he extended a hand towards the light and, smiling, opened it slowly to the gaze of his disciple. A small gold coin shone in the palm.

So these ‘top keyword abridgements’ seem to have something in the way of a narrative progression about them, even a story structure with beginning, complicating middle, highpoint, and a closing if not a resolution. But in finding traces (through a kind of prominent repetition) of narrative structure or story patterning, can the approach be taken beyond traditional, product-oriented text linguistics? How can these abridgements address the question of narrative guidance, of the reader’s experience of and through the text as guiding their expectations of what will happen next, and what will happen finally?
One way in which I would suggest the two issues connect is in the very fact that top keywords – perhaps especially personal names that stand out orthographically due to their initial capitalisation, but all top keywords, by virtue of their disproportionate frequency – become a kind of signposting carrying the reader forward along the way of the story. But what are they signposts of? What do they tell us? My tentative answer is that, like much signage in our lives, they are relatively uninformative – at least, they do not in themselves carry transforming new information. Rather they are reminders, reassurances, that we are on the right road, like route signs that tell us that we are still on the Pennine Way, or the A14. They are more like our ‘straight ahead for X-town’ signs than our ‘turn off here for X-town’ signs. And just as reaching and passing each of the footpath signs that confirms you are on the Heart of England trail is hardly the core or deepest meaningful experience of walking that trail, so the repeated encounters with the top keyword, even in each of its sentential settings, may contribute to the shaping of the reading experience and yet be only indirectly related, even tangential or peripheral, to the deepest experiential effect.

It may also be objected that no reader at the beginning of “Two Gallants” can possibly know that Corley is the top keyword of the story (and even for the second- or third-time reader of the story, this will only be known subliminally). Mutatis mutandis, the same applies to Jill in “Boxes”, and blind in “Cathedral”. That is true enough, and it prompts the question: how far through “Two Gallants” will or must readers get before they can hope to have any awareness that Corley is top keyword, or, reading through “Boxes”, that Jill is top keyword? In one sense this seems a preposterously hypothetical question, but only in one sense; looking at matters in a different light the question is not at all fanciful. It goes along with the idea that by virtue of having just read “Two Gallants” a person has however fleetingly attended to all 3,900 words, in sequence, of that story. They will have made some kind of mental note of them, all 3,900 of them, if they have read all of the text. They won’t be able to recall them, they may not recognise phrases or words from the story even after many readings, but those are different tasks. I have read “Two Gallants” umpteen times, but was still surprised to learn something about Lenehan’s shoes on a recent reading – that they are made of white rubber – which I had apparently never consciously noted previously. And yet I had read and therefore in some sense processed or cognised the relevant information about Lenehan’s shoes on many previous readings – it was not as if this was new text, added to some new, corrected edition of the story.

In normal circumstances, no reader embarking on reading a particular story can know in advance what its top keyword will be. The nearest one might get to ‘prior notification’ would involve the story title: if some word in the title was gradually confirmed as top keyword. In light of this impossibility of advance knowl-
edge of keyword status, it might be some support for a notion of top keywords as foregrounded waymarkers if there was some ‘frontloading’ of the top keyword. By ‘frontloading’ I mean the situation in which what is the top keyword over the course of the whole narrative is disproportionately often used in the opening pages or paragraphs, thereby ‘priming’ the reader on that keyword for the remainder of the story (where it might in fact occur less often). Such a pattern of frontloading certainly applies to the top keyword of “Cathedral”, blind. Blind occurs 80 times in the 6,350 word story, hence with an average frequency of one occurrence every 80 words. But blind is much more prominent that this in the story’s opening lines:

This blind man, an old friend of my wife’s, he was on his way to spend the night. His wife had died. So he was visiting the dead wife’s relatives in Connecticut. He called my wife from his in-laws’. Arrangements were made. He would come by train, a five-hour trip, and my wife would meet him at the station. She hadn't seen him since she worked for him one summer in Seattle ten years ago. But she and the blind man had kept in touch. They made tapes and mailed them back and forth. I wasn't enthusiastic about his visit. He was no one I knew. And his being blind bothered me. My idea of blindness came from the movies. In the movies, the blind moved slowly and never laughed. Sometimes they were led by seeing-eye dogs. A blind man in my house was not something I looked forward to. (emphases added)

In other stories that I have applied top-keyword abridgement to, some degree of comparable ‘frontloading’ is revealed (e.g., in “Boxes”, where the top keyword Jill is disproportionately heavily used in the earlier pages of the story). But in “Two Gallants”, Corley is not frontloaded; on the contrary, if anything it is ‘endloaded’, occurring especially frequently in the final paragraphs. Such evidence again raises questions as to whether the textual dispersal of the top keyword contributes to particular distinct effects. More study is needed to determine when and why frontloading occurs, and when and why it does not.

I have tried to make a case for the limited but genuine importance of top keyword abridgements in our search to understand better the nature of narrative texture, the means by which a narrative guides the reader and nurtures or diminishes expectations, and in our search to find ways in which corpus analyses can be more useful in narrative analysis. We may tentatively conclude that, as an important form of structuring and reader-guidance, a story’s top keyword and its immediate co-texts (the sentences in which that keyword is embedded) are increasingly attended to in the course of reading. The top keyword’s disproportionate recurrence serves to assert and reinforce those sentences’ relevance to the developing action or theme (the situation), as the reading of the story proceeds. The top keywords and their containing sentences are more relevant or central (and more cohesive
and coherent) than much other text; but they may not be the most relevant mate-
rrial, the most defining or expressive of the story’s theme or plot. A top-keyword
abridgement has some of the qualities deemed central to ‘textuality’ (such as inten-
tionality, acceptability, cohesion and coherence, informativeness, interest, inter-
textuality, and situationality) and to narrative prospection. But top keyword sen-
tences are only one element or strand in an integrated network of
progression-signalling elements, dispersed or distributed through the narrative
text. The following chapters explore some of those other key elements.
Keywords and the language of guidance in “The Love of a Good Woman”

1. Story opening as initial guidance

Readers of stories routinely report finding some story developments largely expectable, and their delayed disclosure suspenseful. Certain other developments they will admit they might have foreseen but did not: their occurrence is a surprise, revelation, or anagnorisis. Other developments again are simply deemed ‘possible’ or ‘reasonable’, neither noticeably expected nor unexpected. What sorts of things are guiding readers to such judgements? It is the texture of the narrative text, a complex weaving of strands of meaning, evoked intermittently but in timely fashion so as to create and sustain coherence and continuity. An invaluable general linguistic account of how the lexis in texts evokes and integrates meanings is Halliday and Matthiessen (2004), who call this phenomenon ‘lexicalised ideation’. The present chapter further explores whether those woven phrases and predicates that can be extracted from a story on stylistic, reader-response, and cognitive-schematic grounds can be matched with the lexical prominences derived by corpus linguistic and other semi-mechanical means. Can narratological and reader-ratified judgements about the narrative progression be linked (and, ideally, causally linked) to corpus-derived textual evidence in principled ways?

In this chapter the focus is exclusively on lexical patterning. What are the crucial particularities of the text’s lexical structure that confirm – or divert – the reader’s sense of progression? Can some of this textual guidance be discussed in linguistic terms, in ways that take us beyond motif-spotting and a kind of lexical ‘stamp-collecting’? This chapter explores these issues, and our interest as readers in narrative threads and texture, closure, disclosure, non-disclosure and the implicit, looking in detail at one long story by Alice Munro. In guiding the reader, the narrative text must chart a way between total predictability and an excess of impenetrability. In a sense the modern preference in literary stories is for the text to be, in the course of its reading, partially rather than fully comprehensible. For the literarily competent consumer, stories with a narrativity that is either too obvious or too obscure risk rejection as being unreadable, that is, unreadable in the
sense that no clear or sufficient new benefit appears to be offered in return for the reading effort.

The story which this and the following chapter concentrate on is Munro’s “The Love of a Good Woman”. I look first at the ‘questions’ that readers seem to see the opening of that story posing, for expectable answers in the narrative that follows. The story has a five-paragraph (untitled) opening segment which I will call its ‘Preface’, and this Preface is an excellent test bed for exploring how some portions of text are judged important to the ongoing narrativity, while others are not. Those phrases and sentences in the Preface which are judged most unlikely to contribute to the narrative progression are at least as interesting as the focused-upon material, because they immediately prompt the question why they are retained in the text at all, what purposes they serve, and why it is that readers feel so confident that their purposes are not centrally to do with progression and plot.

Before beginning the analysis, I will summarise the plot, so that the reader is aware of its broad outline. The story is a long one (25,000 words, about 75 pages, comprising the short untitled ‘Preface’ and four titled sections) and contains many characters and episodes, some quite loosely mutually linked. In particular, there are only slight connections between the three pre-teen boys to whom Section I is mainly devoted, and the two women – Jeanette, the dying wife of a farmer (Rupert) and mother of two girls, and Enid, the home nurse who is tending to her – on whom the remainder of the story, Sections II to IV, is centred. At the narrative’s disturbing core is the short Section III: on her deathbed Jeanette reveals to Enid (and to her alone) how an optometrist named Willens made habitual sexual assaults on her during his home visits, and how Rupert finally caught him (or them) in the act, attacked him causing his death, and then conspired with Jeanette to conceal the killing as an accidental drowning. All these disclosures are uncorroborated, and Jeanette dies soon after, so in the final section of the story Enid must weigh up the implications and consequences of what may be lies, misrepresentation, or even fantasy. Here in summary is the story’s plot structure (the four sections and section-titles are Munro’s):

Preface
I. Jutland. Three boys discover the body of Mr Willens in the river, and eventually raise the alarm; his death is presumed to be a tragic accident.
II. Heart Failure. Enid, a “good woman”, attends to the dying Jeanette Quinn.
III. Mistake. On her deathbed, Jeanette discloses to Enid how Willens on numerous occasions forced himself upon her sexually, how on a final occasion Rupert had caught and attacked him, and how they had concealed the killing as an accidental drowning.
IV. Lies. Enid, troubled but attracted to Rupert, prepares to confront him with Jeanette’s confession, but at the close is considering “collaborating in a silence”.

Before beginning the analysis, I will summarise the plot, so that the reader is aware of its broad outline. The story is a long one (25,000 words, about 75 pages, comprising the short untitled ‘Preface’ and four titled sections) and contains many characters and episodes, some quite loosely mutually linked. In particular, there are only slight connections between the three pre-teen boys to whom Section I is mainly devoted, and the two women – Jeanette, the dying wife of a farmer (Rupert) and mother of two girls, and Enid, the home nurse who is tending to her – on whom the remainder of the story, Sections II to IV, is centred. At the narrative’s disturbing core is the short Section III: on her deathbed Jeanette reveals to Enid (and to her alone) how an optometrist named Willens made habitual sexual assaults on her during his home visits, and how Rupert finally caught him (or them) in the act, attacked him causing his death, and then conspired with Jeanette to conceal the killing as an accidental drowning. All these disclosures are uncorroborated, and Jeanette dies soon after, so in the final section of the story Enid must weigh up the implications and consequences of what may be lies, misrepresentation, or even fantasy. Here in summary is the story’s plot structure (the four sections and section-titles are Munro’s):

Preface
I. Jutland. Three boys discover the body of Mr Willens in the river, and eventually raise the alarm; his death is presumed to be a tragic accident.
II. Heart Failure. Enid, a “good woman”, attends to the dying Jeanette Quinn.
III. Mistake. On her deathbed, Jeanette discloses to Enid how Willens on numerous occasions forced himself upon her sexually, how on a final occasion Rupert had caught and attacked him, and how they had concealed the killing as an accidental drowning.
IV. Lies. Enid, troubled but attracted to Rupert, prepares to confront him with Jeanette’s confession, but at the close is considering “collaborating in a silence”.

Before beginning the analysis, I will summarise the plot, so that the reader is aware of its broad outline. The story is a long one (25,000 words, about 75 pages, comprising the short untitled ‘Preface’ and four titled sections) and contains many characters and episodes, some quite loosely mutually linked. In particular, there are only slight connections between the three pre-teen boys to whom Section I is mainly devoted, and the two women – Jeanette, the dying wife of a farmer (Rupert) and mother of two girls, and Enid, the home nurse who is tending to her – on whom the remainder of the story, Sections II to IV, is centred. At the narrative’s disturbing core is the short Section III: on her deathbed Jeanette reveals to Enid (and to her alone) how an optometrist named Willens made habitual sexual assaults on her during his home visits, and how Rupert finally caught him (or them) in the act, attacked him causing his death, and then conspired with Jeanette to conceal the killing as an accidental drowning. All these disclosures are uncorroborated, and Jeanette dies soon after, so in the final section of the story Enid must weigh up the implications and consequences of what may be lies, misrepresentation, or even fantasy. Here in summary is the story’s plot structure (the four sections and section-titles are Munro’s):

Preface
I. Jutland. Three boys discover the body of Mr Willens in the river, and eventually raise the alarm; his death is presumed to be a tragic accident.
II. Heart Failure. Enid, a “good woman”, attends to the dying Jeanette Quinn.
III. Mistake. On her deathbed, Jeanette discloses to Enid how Willens on numerous occasions forced himself upon her sexually, how on a final occasion Rupert had caught and attacked him, and how they had concealed the killing as an accidental drowning.
IV. Lies. Enid, troubled but attracted to Rupert, prepares to confront him with Jeanette’s confession, but at the close is considering “collaborating in a silence”.

Before beginning the analysis, I will summarise the plot, so that the reader is aware of its broad outline. The story is a long one (25,000 words, about 75 pages, comprising the short untitled ‘Preface’ and four titled sections) and contains many characters and episodes, some quite loosely mutually linked. In particular, there are only slight connections between the three pre-teen boys to whom Section I is mainly devoted, and the two women – Jeanette, the dying wife of a farmer (Rupert) and mother of two girls, and Enid, the home nurse who is tending to her – on whom the remainder of the story, Sections II to IV, is centred. At the narrative’s disturbing core is the short Section III: on her deathbed Jeanette reveals to Enid (and to her alone) how an optometrist named Willens made habitual sexual assaults on her during his home visits, and how Rupert finally caught him (or them) in the act, attacked him causing his death, and then conspired with Jeanette to conceal the killing as an accidental drowning. All these disclosures are uncorroborated, and Jeanette dies soon after, so in the final section of the story Enid must weigh up the implications and consequences of what may be lies, misrepresentation, or even fantasy. Here in summary is the story’s plot structure (the four sections and section-titles are Munro’s):

Preface
I. Jutland. Three boys discover the body of Mr Willens in the river, and eventually raise the alarm; his death is presumed to be a tragic accident.
II. Heart Failure. Enid, a “good woman”, attends to the dying Jeanette Quinn.
III. Mistake. On her deathbed, Jeanette discloses to Enid how Willens on numerous occasions forced himself upon her sexually, how on a final occasion Rupert had caught and attacked him, and how they had concealed the killing as an accidental drowning.
IV. Lies. Enid, troubled but attracted to Rupert, prepares to confront him with Jeanette’s confession, but at the close is considering “collaborating in a silence”.
Chapter 5. Keywords and the language of guidance in “The Love of a Good Woman”

After Jeanette’s death early in the final section, Enid decides to question Rupert (whom she has hopes of one day marrying) about Jeanette’s deathbed confession. At least, she devises a melodramatic scheme in which she will confront him during a boating excursion on the local river and, if the shame and punishment of going to the police is too much for him then – she thinks – he can drown her. But even as she embarks upon this plot she begins to have second thoughts, and thinks how “collaborating in a silence” or a secret can bring benefits and “keep the world habitable”. The final section is a remarkable one, full of irrealis hypotheticality, of what Enid thinks will happen and what she would do in those circumstances. The prominence of would in that final section is reflected in a keyword analysis: would is the seventh-ranked keyword in that section, with 50 instances and, at 0.94% of the text, is more than three times as frequently used as in the reference corpus (0.28%). Although would is not discussed further in the present chapter, since the focus here is on fully lexical keywords in the story, modals like these are discussed at length in Chapter 8.

The story ends before we discover whether Enid actually confronts Rupert or not. No mid-river confrontation is narrated, and judging by the story’s brief proleptic Preface, perhaps none took place. The most blatant plotting in the story then is Enid’s, not Munro’s, and in fact we never know how much of it is enacted. But we try to guess, of course. We have formed some expectations. We may clutch, for example, at textual scraps like the following, very close to the end of the story, narrated from Enid’s point of view when Rupert goes into a woodshed to find her some boots to wear on the walk to the lake:

A house like this, lived in by one family for so long a time, and neglected for the past several years, would have plenty of bins, drawers, shelves, suitcases, trunks, crawl spaces full of things that it would be up to Enid to sort out, saving and labeling some, restoring some to use, sending others by the boxload to the dump. (Munro 1999: 76)

That may warrant us in surmising that at some point after the boating excursion Enid finds the box of optometry instruments in the shed, hidden there by Jeanette, and becomes the anonymous dispatcher of them to the museum. The impersonal museum caption (in the second paragraph of the story Preface, reproduced below) could hardly be worded as it is if it had become public knowledge that Mr Willens had not died by drowning; so presumably Rupert did not go to the police. But beyond this we approach secrecy: Enid may or may not have confronted Rupert over Willens’s death. He may or may not have admitted a part in it (he may have been entirely innocent, of course), and he and Enid may or may not have married. These episodes are part of a different narrative, not the one told, and so are not available to be experienced.
Using several small groups of informants, I have made preliminary attempts to use readers’ responses to explore the story’s texture more systematically. For these purposes, I focused on the story’s brief Preface- or prologue-like opening five paragraphs, a one-page title-less text. I have shown that Preface to several dozen literate, adult readers, canvassing their responses via questionnaire and follow-up discussion, but I make no claims for the statistical significance or reliability of their comments. Thus I have removed the Preface text from the story which it opens, but have told readers that what follows is a long story, and have then asked them to respond to it as a story opening. Here, then, is that story Preface:

For the last couple of decades, there has been a museum in Walley, dedicated to preserving photos and butter churns and horse harnesses and an old dentist’s chair and a cumbersome apple peeler and such curiosities as the pretty little porcelain-and-glass insulators that were used on telegraph poles.

Also there is a red box, which has the letters “D. M. Willens, Optometrist,” printed on it, and a note beside it, saying, “This box of optometrist’s instruments though not very old has considerable local significance, since it belonged to Mr. D. M. Willens, who drowned in the Peregrine River, 1951. It escaped the catastrophe and was found, presumably by the anonymous donor, who dispatched it to be a feature of our collection.”

The ophthalmoscope could make you think of a snowman. The top part, that is – the part that’s fastened onto the hollow handle. A large disk, with a smaller disk on top. In the large disk a hole to look through, as the various lenses are moved. The handle is heavy because the batteries are still inside. If you took the batteries out and put in the rod that is provided, with a disk on either end, you could plug in an electric cord. But it might have been necessary to use the instrument in places where there wasn’t any electricity.

The retinoscope looks more complicated. Underneath the round forehead clamp is something like an elf’s head, with a round flat face and a pointed metal cap. This is tilted at a forty-five-degree angle to a slim column, and out of the top of the column a tiny light is supposed to shine. The flat face is made of glass and is a dark sort of mirror.

Everything is black, but that is only paint. In some places where the optometrist’s hand must have rubbed most often, the paint has disappeared and you can see a patch of shiny silver metal.

And here is a version of the main question I then asked these readers:

Are there any words or sentences in this story opening that you feel or suspect are especially relevant to how the story proceeds, or what is likely to happen later? If there are, please underline or highlight them. As a rough guideline, you might pick out 4 or 5 clauses, of 10 to 20 words in total.
The words and phrases most frequently singled out by respondents were: *escape... catastrophe, Mr Willens, drowned, red box,* and *donor.* Invited to comment on why they had highlighted just the words and phrases they had, nearly all the readers reported that they expected the subsequent story to explain the ‘catastrophe,’ and to tell of how Mr Willens came to drown in 1951. Some respondents linked the instruments’ escaping the catastrophe with the idea at the end of Paragraph 3 that Mr Willens might have made rural home visits to places without electricity, and began to sketch a scenario in which he took his instruments on a home visit, but somehow became separated from them so that he drowned in the river but his equipment did not. They also assumed that the equipment must have some bearing on subsequently reported events, to justify the long dwelling upon them here. A few respondents also wondered why the instruments had become museum pieces, unless they were in some sense relics of a remote way of doing things, a now-distant way of life. After all, if a dentist or general practitioner died suddenly today, one would not normally send the tools of their trade off to the local museum for display. The optometry equipment that is described was presumably not archaic or historically interesting back in 1951, so something seemed to need further explaining, those readers felt. Some also found curious, and therefore noteworthy, the caption’s mention of instruments ‘escaping the catastrophe’ (why should not everything of Mr Willens’s, save the clothes he was wearing, have escaped the catastrophe?) and its speculation that the donor was also the finder of these instruments (implying they were lost or mislaid). The anonymity of the donor, too, was felt to be mysterious – not because anonymous donation to museums is out of the ordinary, but because how or why it should arise in relation to a box of optometry instruments was hard to imagine. The reader of the full story learns that Mr Willens has left a widow, and may wonder why the ‘found’ instruments were not returned to her rather than the museum; but at the story opening we do not yet know that a Mrs Willens exists.

Despite the descriptive tone of the Preface, many respondents found the optometry instruments hard to visualise; some even wondered whether one was expected to try. Very few words from the two main paragraphs describing the instruments were underlined by readers, as likely to be elaborated upon in the story. Although several colloquial moments in the description were commented upon, one sentence was particularly cited: “Everything is black, but that is only paint.” Together with the following sentence, this made some readers think about surface veneers concealing a substance or depth of a different kind, and how appearances can differ from underlying reality.

What is reader-guidingly significant in the story’s prologue, and how can one justify selection of certain words or propositions as ‘most significant’? An answer to that question, it seems, must invoke cultural, psychological and genre-based
criteria. In the present case, the demands of narrativity prompt us particularly to attend to significant events (as distinct from states or descriptions). And as Labov (1972) like others long ago noted, as an event of human significance, nothing beats death or danger of death. On the above bases, then, arguably the most significant event alluded to in the opening passage from “The Love of a Good Woman” is the following:

Mr. D. M. Willens [who] drowned in the Peregrine River

This phrase was singled out by most of my informants as a likely ground for subsequent narrative elaboration. As an event-report, it contrasts with much of the rest of the Preface, which largely comprises reported states (there is a museum, there are the following optometrist’s instruments in that museum, the instruments include an x, which has features a, b, c, etc). This is all detachedly reported, at least until the final, more evaluative paragraph, which as noted memorably begins Everything is black, but that is only paint.

In context, that sentence is striking and memorable for several reasons: its generality, its referential vagueness or indefiniteness, and its nearness to rhetorical self-contradiction. The generality stems from the use as subject of everything, one of those indefinite pronouns (discussed more fully in Toolan, forthcoming) often found in situations of significant narrative gap or vagueness, which are most narratively important when they fill an entire phrasal construction (e.g. Everything was black rather than Everything she was wearing was black). The sentence Everything is black, but that is only paint is also memorable for reasons to do with its apparent generalisability and potential figurative interpretation. Rhetorically it attracts our attention by means of its hyperbole. It is obviously absurd to say that ‘everything’ is black. It then presents a self-refutation: nothing is really black, only the painted surfaces are. The sentence also achieves a slight informality and succinctness, after the unmotivated specificity of the preceding paragraphs, with their perhaps deliberately-laboured descriptions of the instruments, and the (again, perhaps intentionally) doubtfully apposite analogies (snowman, elf’s head).

Because of these particular child-oriented analogies, and the phrasings in which they are embedded, parts of these paragraphs are reminiscent of the explanatory language of a teacher, taking their pupils around the museum, using the language of modalised negotiation rather than categorical description (“What does this make you think of? Yes, the top part does look like a snowman, doesn’t it. Feel how heavy the handle is…”). But the discourse effects a change of plane or register with Everything is black, but that is only paint. I am therefore inclined to treat it as a key evaluative guide, alongside Willens drowning in the Peregrine River as a key referential guide, for what follows. Here I am using Labov’s (1972) fundamental distinction, in his analyses of oral personal narratives, between referential clauses
(telling the events: what happened) and evaluative clauses (conveying the teller's reactions to what happened, and the point of the story. See Labov 1972, Chapter 9; Toolan 2001: 143–147).

The interim conclusion then is that the following clauses (and their constituent words and phrases) in the Preface seem likely to be the most noticeably relevant to the ongoing narrative:

Mr. D. M. Willens, who drowned in the Peregrine River, 1951… there is a red box [which] escaped the catastrophe and was found, presumably by the anonymous donor,

and perhaps the entirety of the final paragraph of the Preface:

Everything is black, but that is only paint. In some places where the optometrist's hand must have rubbed most often, the paint has disappeared and you can see a patch of shiny silver metal.

How satisfactorily principled or rule-governed is this selection? The more we can identify principles underlying the postulated selective attention, the less vulnerable is the whole exercise to dismissal as purely arbitrary and subjective.

A first rebuttable presumption that it may be useful to make about narrative paragraphs is that their first main clause (or clauses, in the case of initial compound clauses) is a position of prominence. This presumption reflects the special status accorded in composition theory to the first sentence in any well-formed paragraph, and the special naming of it as the topic sentence. If the presumption is correct, paragraph-initial main clauses, other things being equal, are likely to bear narratively important information and should be especially attended to. On that basis the following clauses in the preface would be selected:

For the last couple of decades, there has been a museum in Walley
Also there is a red box
The ophthalmoscope could make you think of a snowman.
The retinoscope looks more complicated.
Everything is black, but that is only paint.

At this point, since narrative progression or prospection is the consideration, some or even all of these initial main clauses might be set aside after this initial selection. What makes this reader, at least, discard the first of these as chiefly situational, and the third and fourth as non-prospectively specific, while the second and fifth both feel potentially prospective, even though one is specific and the other more general? There seems to be a sense of scale or proportion that is applied here, and indeed enacted throughout the Preface: museum denotes too general and inclusive a category, butter churn and retinoscope are too particular a level of reference, in
the textual circumstances. That is not to deny that a whole story could be about a museum, and take the museum as main ‘protagonist’, or that other stories could do the same with a butter churn or retinoscope.

But this particular Preface mentions a human being, and his unnatural death, and some slightly curious details surrounding that death, all in the latter half of Paragraph 2. Such ‘human interest’ information, other things being equal, can be presumed to take precedence over (and conditions our assessment of) the otherwise mildly foregrounded mentions of the museum and the ophthalmoscope. Thereafter, lexicalised items which are potentially agentive, instrumental, or force-like (potential ‘doers’ in transitivity terms), rather than those more naturally interpretable as circumstantial or to do with setting, attract greater attention. Arguably they tend to be notionally at a comparable level of wholeness or compositionality as is the named individual, Willens, and perhaps bear on his character directly rather than relating chiefly to the equipment he used (information which is too specific to be likely to be central in the story), or the purposes of the Walley museum (information which is too general). On these bases the mentioned existence of the red box (but not all the detail about its content) remains salient, as does the final paragraph about the black painted surface of the instruments, rubbed away in places to reveal shiny silver. Because there is an underlying pressure on narratives to be particular, those sentences in which markedly general statements occur or where what I am calling complex wholes not related to setting or circumstances (potential agents and instruments) are named, tend to be thrown into relief. An example of the markedly general statement is Everything is black, but that is only paint; examples of ‘instrumental’ complex wholes are Mr D. M. Willens and the red box. They are noticeable and intended to be noticed, albeit for different reasons. Reasons relating to potentially agentive complex wholes have been outlined above. As for generalising sentences, noticeability arises because in the context of a particularising narrative, the generic or quasi-generic sentence like Everything is black…in principle has no place: the generic sentence is by virtue of its claimed indiscriminate truth an ‘anti-narrative’ sentence (Toolan 1998: 64). In both cases Gricean relevance can be assumed to operate: a teller’s repeated mention of any potentially-agentive entity and inclusion of any generic statement implies that both are especially relevant, to justify the implicit invitation to the reader to add them to their picturing of the situation.

I believe that a reader proceeds through a narrative text carrying forward maximally vague mental pictures of the situation projected by that text (a view defended in Toolan forthcoming). On that basis I will here hypothesise that a reader advances, and is supposed to advance, beyond the Preface with a vague mental picturing which comprises: a Mr Willens drowning in the river, his red box of instruments escaping the catastrophe, their unspecified finder/donor, and the
black surfaces of those instruments sometimes rubbed away to the shiny silver beneath. All this amounts to two or three maximally vague images, with only a very few labels attached:

- Willens, drowned, River, red box, instruments, escape, anonymous donor, optometrist’s hand, black, and shiny.

These images and lexical labels are the resources with which a reader can formulate propositions that recall key parts of their impression of the story thus far. Generally, readers recall the gist and not the precise syntagmatic wordings, so that, for example, some readers, if questioned, may refer to the instruments escaping the tragedy rather than the catastrophe. By the same principle, hundreds of words, concepts and visualised descriptions or reports, directly derivable from the Preface, are not reliably or particularly relevantly carried forward at all: words such as butter churns, museum, snowman, paint, and batteries.

Among the items just cited, the word and concept paint is an interesting case ("that is only paint"), since an important secondary role emerges for paint in Jeanette's Section III confessions: she tells Enid she has concealed the bloodstains in the parlour by covering the floorboards with a noxious-smelling brown paint, which she then associates causally with her illness. But whether a reader leaves the Preface expecting paint to be instrumental in the narrative progression seems unlikely, given the assumptions sketched above: our according narrative priority, normally, to humans, potentially agentive entities, and the repeatedly-mentioned lexical items. The details of Willens's drowning and the retrieval and special treatment of his optometry instruments are far from the only 'unknowns' alluded to but not resolved in the Preface; but it seems reasonable to argue that the reader can draw on textual signalling and world knowledge to assume (perhaps wrongly) that these unknowns will be most central to the narrative progression, and other unknowns will not. Thus the Preface goes to the trouble of speculating that the mains-powered retinoscope retained the option of battery-powered operation since the device might have needed to be used in visits to rural homes off the electrical grid. But this supposition is not a particularly salient one for the story that follows, nor is it explicitly confirmed or returned to in the later text.

Not only do we read on with an extremely vague picture of Mr Willens drowning, we do so also aware that the picture is unsatisfactorily vague and in need of focussing and detailing; we seek such clarification in the text ahead. We also carry forward a vague picture of his instruments escaping the catastrophe, and also – and this is surely the hardest picture to begin to visualise – a vague picture of this box of ophthalmology instruments 'being important' in some scene or imaginable episode. If one goes to a great deal of relevance-theoretical effort, perhaps one concludes that this latter imaginable but necessarily entirely undefined scene is
likely to be one involving or arising out of an eye-examination, and one that most likely involves Mr Willens and a patient.

One model of text structure proposes that a text comprises a covert sequence of segments, each new segment in turn ‘answering’ the question(s) raised by the preceding ones (see, e.g., Hoey 2001). If modern short stories like Alice Munro’s can be assimilated to such a model at all, then it seems that the question-and-answer dialogues they imply are often of an exceptionally ‘open’ or liberal kind. Any structuring by means of implicit prospecting questions and their prospected answers seems to be more attenuated and dispersed in literary stories than in non-literary texts. This is because in literary text many questions are triggered by each segment rather than just a few or one; and the ‘answers’ are sometimes postponed or withheld (as occurs with narrative gaps). Nevertheless, as I have assumed in eliciting readers’ expectations prompted by the Preface, a story’s opening passage can be treated as posing questions for subsequent textual answering. Just what those questions are can be highlighted or reserved on the basis of relatively stable criteria such as inherent eventhood, human interest, gravity, generality of claim, textual positioning or emphasis (especially paragraph-initial position), and consonance with received cultural schemas. Another way to consider this issue would be to suggest that the Preface raises nothing as narratively tellable other than the death by drowning of Mr Willens and the remarked-upon ‘escaping of the catastrophe’ of his optometry instruments. Thereafter, Section I tells us how Willens’s body was discovered, inside his car by three pre-teen boys who have headed out from Walley so as to be the first to take an icy spring dip in the Peregrine River. The section goes on to tell a great deal more about the boys (and their families) and how they delayed telling of their exciting discovery. But the Preface and Section I together in turn leave several reasonably-prompted questions awaiting answers: How (if accidental) or why (if deliberate) did Willens come to drive his car into the river? And – in itself less importantly – was the fact that his instruments were apparently not with him accidental or planned? If those are the major narrative questions carried over to Section II, it soon becomes apparent that they are almost entirely neglected through the course of the latter section which, as noted earlier, focuses entirely on Enid and the dying Jeanette.

Of course the images and labels singled out above from the Preface are not the only things a reader ‘carries forward’: there is also the question of what our cultural and real-world knowledge, our background scripts and generic expectations, brought to bear on this story opening, lead us to assume and expect. What are our scripts, mental models or schemas, for a D. M. Willens drowning in the Peregrine River? What words, phrases, even full sentences, does this event or episode license and make expectable, or likely, or relevant, at some point in the remainder of the story-text? Among the words and phrases that ‘Willens drowned in the Peregrine
River’ predicts or makes expectable for me, on the basis of my real-world cultural schema for human drowning are the following: death, water, accident, current, suicide, out of his depth, overpowered, weak(ened), unconscious, and many more. Similarly licensed or predictable, among sentences, would be the following: It was an appalling tragedy, No-one could help him, No-one saw how it happened, The river was exceptionally high, The river was dangerous, among others. These example words, phrases and clauses are just the tip of the iceberg of schematised predictability, for a competent reader of the foregrounded ‘Willens drowned’ event-report. While these are now possible and perhaps even primed for occurrence in the ensuing story, the next section considers which words are actually prominent in the text as a whole, using Keywords as an objective means of searching.

2. Top lexical keywords as narrativity indices

Can some of the notions (words, propositions, and questions) singled out above now be matched with those that might emerge from corpus-analytic methods, such as Keywords, and collocational analysis? Use of Keywords in a study so essentially syntagmatic and developmental as this one is fraught with difficulties and limitations; but these are best seen by making the attempt. Nor is it simply a matter of conceding that keywords may not bear on plot-progression because they have more to do with thematic preoccupations. There is no direct link between a text’s keywords and its themes, structure, or point (however these may be defined), any more than there is with plot. But given that literary word-choice is often crucial to the particularity of effect and is made with extreme care, it seems reasonable to expect that there may be indirect links with both theme and plot.

A beginning was made on the identification of keywords by comparing lexical frequencies in the Munro story with frequencies in my half-million word comparator corpus of modern fiction, prepared for purposes of close comparison. Unlike the use of Keywords on “Two Gallants” and other stories in the previous chapter, where the top keyword became the basis for an abridgement, here I elected to focus exclusively on the lexical and semantic. Accordingly I generally excluded grammatical items from further attention (e.g., frequencies of numerous grammatical words such as the, in, of, had, etc.), and similarly set aside proper names and place names. I did the latter on the assumption – certainly a challengeable one – that it is not informatively revealing to find confirmation that names like Enid, Rupert, Walley, Willens etc. are enormously more frequent in the story than in narrative fiction generally. The table below lists the story’s keywords once most grammatical items and obvious proper names have been removed. As the rankings in Column 1 indicate, only one of the top thirteen keywords, by the log likelihood
### Table 3. Positive lexical keyword listing for “The Love of a Good Woman”

<table>
<thead>
<tr>
<th>N</th>
<th>WORD</th>
<th>Freq. in LGW and %</th>
<th>Freq. in Reference</th>
<th>Corpus and %</th>
<th>Keyness</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>car</td>
<td>40 0.16</td>
<td>57 0.01</td>
<td>108.0</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>mother</td>
<td>72 0.28</td>
<td>319 0.07</td>
<td>81.7</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>said</td>
<td>232 0.91</td>
<td>2,400 0.54</td>
<td>50.2</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>water</td>
<td>42 0.16</td>
<td>185 0.04</td>
<td>48.0</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>riverbank</td>
<td>7 0.03</td>
<td>0 0.00</td>
<td>40.8</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>children</td>
<td>35 0.14</td>
<td>157 0.04</td>
<td>39.1</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>front</td>
<td>33 0.13</td>
<td>145 0.03</td>
<td>37.8</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>forsythia</td>
<td>6 0.02</td>
<td>0 0.00</td>
<td>35.0</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>phoned</td>
<td>6 0.02</td>
<td>0 0.00</td>
<td>35.0</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>smell</td>
<td>19 0.07</td>
<td>53 0.01</td>
<td>33.5</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>nursing</td>
<td>8 0.03</td>
<td>4 0.00</td>
<td>31.8</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>boys</td>
<td>24 0.09</td>
<td>98 0.02</td>
<td>29.8</td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>jail</td>
<td>6 0.02</td>
<td>1 0.00</td>
<td>29.3</td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>pills</td>
<td>5 0.02</td>
<td>0 0.00</td>
<td>29.1</td>
<td></td>
</tr>
</tbody>
</table>

Calculation, is fully lexical: *car*. The items filling the first to eighth keyword rankings, here deleted, are: *Enid, Rupert, Quinn, Willens, or, Cece, they*, and *Bud*. What follows is, however, a ‘full’ listing of positive lexical keywords, in the sense that *pills* is the last identified lexical keyword by log likelihood means.

In the analysis that follows, there is a real difficulty as to whether or not the item *mother* should be excluded from consideration on the grounds that it is chiefly used in the naming of several women characters where proper names are not used: Cece’s mother, Bud’s mother, Enid’s mother, and so on. On the other hand, notwithstanding its naming function, it remains a lexical item, and the only such kin term to be so prominent as to be key, so there are equally grounds for continuing to include it in the analysis. *Said*, also, might be felt not to merit further consideration in relation to narrative progression, as it is chiefly a reflex of the numerous occasions of speech reporting in the story.

Besides these two items, the most prominent fully lexical keywords that emerge are *car, water, riverbank, children, and front*. But even this short list includes one low-frequency item (*riverbank*, which occurs just seven times and comprises .03% of the text), and the last two items listed (*children and front*) have a keyness score below 40. The list also includes one item, *front*, whose relative high frequency may be simply a quirk of Alice Munro’s style, in many of her stories (as, perhaps, is her comparatively high use of *or*, the fifth-ranking keyword in the story and, at 0.76% of the text, three times as frequent here as in the reference corpus).
Nevertheless we can proceed with *water*, which is four times more frequent (at 0.16% of the text) than in the specially-prepared fiction reference corpus (0.04%), and seven times more frequent than in the Cobuild Bank of English’s US Books corpus. As it happens, *water* appears no fewer than five times in the opening paragraphs of Section 1, that is, in the text that immediately follows the ‘Preface’ quoted earlier. Nearly all these uses refer to the boy’s finding of the car in the water (and then, on closer inspection, of Mr Willens’s body inside the car): *car tracks to the water’s edge…they would jump into the water…pale-blue shine to the water*. It is possible, then, that *water* is important in guiding the reader here, given the ‘induction’ that the Preface (which does not use the word *water*) has provided.

A focus on WordSmith-identified keywords is no more than a focus on the most disproportionately repeated words of a text (with, here, the decision to exclude grammatical items, proper names, etc.). But how much will evidence of ‘noticeable’ or ‘heavy’ use of a word like *water* or *car* in this story tell us? Among the things it does not cover are development, variation and change, where a word used at one point is in effect displaced by use of a cognate or variant term later on. Counting repetitions very emphatically neglects such change and modulation, which may be important in the process and progression of narratives as in other texts. So besides repetitions we have to consider all the associations, collocates, semantically-related terms, that are arguably activated or at least made more accessible to activation, by the use of a particular ‘starting’ term.

Consider *water* again: besides all its collocates, its semantic associates include *drought, moist, earth, air, fire, land*, and many more lexical items, which are neither collocates nor synonyms. Its twenty most frequent lexical collocates as recorded in the Cobuild collocates CD are, in order of frequency: *hot, cold, under, food, drinking, no, like, boiling, down, through, electricity, supply, some, off, supplies, running, warm, fresh, air, and deep*.

But dwelling on disproportionately frequent words and their collocates and semantic associates is a focussing on the ‘what is said’ rather than the unsaid but implicit, derivable by implicature on the basis of what is said and the assumed context of utterance. For the latter purposes, some version of Gricean implicature must be applied to our reading of narrative, and it must often be the case that such implicatures, in combination with the stated message, are what guide the reader. It is unclear how effectively corpus analytic methods can assist in highlighting such implicit cues, and this is a central issue throughout this study. If the reader’s experience of narrative progression entails an integration of the explicit and the implicit, and if corpus analyses only uncover trends within the explicit, how sufficient (or incomplete) is such a study of (putative) explicit features of progression or narrativity? Would such a corpus – or any stylistic – study be radically incomplete, or is the ratio of explicit to implicit indications weighted towards the explicit, with
the implicit material sufficiently derivable from the explicit that an analysis narrowly focussed on the surface form of the text remains valid? My working assumptions throughout this study are that there are textual signals of those implicatures or inferences that are important to the unfolding narrative, and that some of them can be identified by combinations of partially generalised means.

Let us return to the Preface, now, where it was argued that drowned was key to progression expectations. Consultation of the Cobuild Collocations CD reveals that water and river are among the top twenty collocates of drowned, as node word, in the Bank of English corpus. Therefore it is reasonable to suppose that anyone reasonably fluent in English who encounters the word drowned in coherent discourse will have the tacit expectation that the words water and river may well appear in the vicinity, in prior or subsequent text. The expectation is arguably stronger where drowned has been used literally rather than metaphorically (although the Collocations dictionary does not distinguish these). A strong collocational relationship is not one of entailment; nevertheless it is a kind of lexical-prosodic predisposition which means that finding the word water soon after one has read the word drowned is no surprise and strongly foreseeable. And it is important that the priming word in the prefatory section is drowned, which strongly predicts water proximately, and not water, which, on its own, does not predict drowned at all.

Towards the end of the story, water is frequent in two paragraphs of Enid’s speculative narrative about confronting Rupert mid-river, as are several words semantically closely associated with water, such as river and swim. Again, such claimed association is a matter of judgement rather than objectively provable, and where one chooses to draw the line between ‘closely associated’ and not closely associated is debatable (e.g., is the word boat that occurs in the passage a semantic associate of water?). But lexico-semantic clustering around water, river, swim, sink, and shore is so prominent that it is reasonable to argue that it has to be prominent, too, in a reader’s experience, their grasp of what is centrally evaluated and tellable as the story’s chief preoccupation at this point.

3. Textual segmentation and keywords’ collocates

In a first effort to break free from the totalising perspective, and to be more sensitive to the incremental development of the text, the story was subdivided into five parts for keyword analysis: the brief introductory passage and the four main sections. Wordlists were drawn up for each of these five parts, and on the basis of comparison with the reference corpus of fiction, those keywords that occurred at least four times in a section were identified. Again, proper names and most (but not all) grammatical items were excluded. One or two interesting patterns emerged (see Figure 1).
These keywords do reflect some of the narrative preoccupations, if not the narrative progressions, of the story. Take the word *mother*, for instance, key in Sections I and II. Mothers are quite prominent in the first long section, even though it is ostensibly about the three boys and their finding of Mr Willens and his car in the river. All three of the boys’ mothers are described at some length, suggesting that each is (still) the most important figure in her son’s life. This story-length section might be interpreted as in dialogue with some of Munro’s earlier ‘lives of girls and women’ stories; here we meet the ‘lives of boys and mothers’. In Section II, Heart Failure, a fourth mother looms quite large, though we really only see her indirectly; this is Edith’s mother, with whom Edith lives and on whom she is still in some respects materially and psychologically dependent. But *mother* as a term and as a role drops away in the latter half of the story when, particularly in light of Jeanette Quinn’s disturbing revelations, Edith ceases to be mothered and devises her own strange plan of action. She also, as it happens, becomes something of a mother herself, to Jeanette’s two little girls.

A further expansion of the keywords picture in Figure 1 can be effected by adding to the above array all lexical items, independently identified as being among the commonest collocates of the keywords, which are in addition relevant in that they occur in the text: the results are presented in Figure 2 below. For this purpose I consulted the Collins Cobuild Collocations CD for the top twenty collocates of each of the identified keywords, where the latter were themselves sufficiently core vocabulary to have their collocates recorded on the CD. With reference to Section I, for example, among *car’s* top twenty collocates, one (*police*) is particularly relevant and actually occurs several times in the text, so this is duly recorded. None of the top recorded collocates of *boys* occurs in the story, while *forsythia* is itself too ‘non-core’ a vocabulary item to have its collocates (if it has any) recorded in the CD.
<table>
<thead>
<tr>
<th>disk</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>car</strong> (police), <strong>boys</strong> (girls, school, old, men), <strong>forsythia</strong>, <strong>police</strong> (officers, told, secret, car), <strong>eggs</strong>, <strong>mother</strong> (father, children, home), <strong>phoned</strong> (police, home, hospital, mother), <strong>flats</strong> (block, houses, mud), <strong>home</strong> (back, come, family, came, went) <strong>block</strong> (flats, road, office)</td>
</tr>
</tbody>
</table>

| mother (father, child, daughter, old, children, home), **promise** (made, keep, new, kept), **nursing** (home, care, private, patients, hospital), **pills** (sleeping, take, taking), **kidneys** (heart), **wanted** (said, know, just, go, people), **hospital** (taken, patients, bed, doctors) |

| car (back, got, down, drove), **cloth** (table, wrapped), **glug, got** (know, just, people), **up, just** (know, people), **box** (see), **leg** (first, right) |

| water (hot, drinking, down, deep), **boat** (people, small, trip, down, water), **jail** (years, sentence, year, month, go, time), **river** (bank, water, side), **punished** (caught, law) |

**Figure 2.** A keywords map for “The Love of a Good Woman”

Figure 2 lists only the lexical keywords of the sections (with inclusion of a very few grammatical-lexical borderline items, such as up and just). One noticeable feature is that with each successive main section the list of keywords shortens, from ten items in Section I to five in Section IV. This is by no means a natural trend in multi-section narratives, and is possibly indicative of an increasing thematic focussing in the course of the story. The keywords also seem increasingly to group into a few sets. That judgement, it should however be acknowledged, may well be influenced by prior familiarity with the story text. Thus it is reasonable to see the five keywords of Section IV as falling without remainder into two sets: jail and punished ‘go together’, as do water, boat and river. All the lexical keywords of Section III relate directly to ‘concealing Willens’s murder’, and none, interestingly, derive from his reported assaults on Jeanette. But to see this, prior acquaintance with the text is clearly required. Among the keywords of Section II, nursing, pills, kidneys and hospital – about half of them – arguably form a set. Grouping of the more numerous keywords in Section I into sets would be contentious and perhaps unjustifiable.

Using the Cobuild collocations CD trusts in the reliability of this resource, but there are some grounds for concern that its database is not a fully representative sample of mainstream English, despite being based on a huge corpus. For example, the sixth and eighth most frequent collocates of leg are identified as spinner and before, with suspiciously many cricketing examples of leg spinners and leg before wicket. For the purposes of identifying a word’s commonest collocates there are now many electronic and book-format resources such as the online Cobuild Concordance and Collocations Sampler; or Kjellmer (1994), but for ease of on-screen compilation and cross-referencing the Collins Cobuild CD was particularly convenient. Interestingly the online Cobuild collocations sampler, using a 56
million word corpus, gives *spinner* as only the fifteenth ranked collocate of *leg* (using the t-score calculation: t-score is a statistical measure which adjusts for the overall frequency of a word, when calculating the confidence with which we can predict that two words co-occur more often than chance).

In addition, the listing of occurring common lexical collocates here, in parentheses, after identified lexical keywords, should not be over-interpreted. Thus, with reference to Section IV: *Lies*, after the keyword *jail*, one of its Cobuild-identified commonest collocates *sentence* is listed simply because (and Wordlist confirms) *sentence* occurs once somewhere in that same 5,300 word section: there is no guarantee that *sentence* is used in close proximity to *jail*, or even that it is used in the ‘prison term’ rather than ‘unit of writing’ sense. As it happens, however, a brief check reveals that the single occurrence of *sentence* is used in the ‘prison term’ sense, and just three lines after mentions of *jail* (bold added):

> Every day, or as often as they will let her, she will sit and talk to him in jail, and she will write him letters as well. If they take him to another jail she will go there; even if she is allowed to see him only once a month she will be close by. And in court—yes, every day in court, she will be sitting where he can see her.

> She does not think anyone would get a death *sentence* for this sort of murder…

(Munro 1999: 74)

4. Local (within-section) interrelation and collocation of keywords

What does Figure 2 above show? To begin with, it confirms the impression, drawn from the keyword listings in Figure 1, that many of the keywords in a particular section are directly or indirectly related to each other. Thus in Section IV, *water*, the top lexical keyword, is confirmed as a collocate of the two other keywords, *boat* and *river*, with which it was grouped earlier on the basis of introspection. Or, in Section I, where direct links among the lexical keywords viewed on their own were least apparent, the range of occurring collocates helps to articulate the likely links between the keywords *home*, *mother*, and *boys* (and, more tangentially) *phoned*. This is because *family* is recorded as an occurring collocate of *home*, while *home* itself is a collocate of the keyword *mother*, as is *children*. Taking this group of keyword or collocating items together, an indirect link to the second most key keyword *boys* becomes more likely, drawing on real-world knowledge to produce a schema-like ‘definition’ comprising nine of these lexical items (in bold): *The boys (and girls) are/were children, with a mother and father, in a home, which they came and went back to*.

By this means, the word-listings for individual sections now appear rather more coherent, and more amenable to some attempt at a narrativising synthesis. Thus
the occurring collocates listed for Section II give greater encouragement to the speculation that a promise has been made and kept, that the promise was made by and to two among the following parties: mother, father, daughter, patients, people, and doctors; that some of the latter (particularly patients) unsurprisingly, are being nursed in bed at home or in hospital (and not just by the mentioned doctors, since they do not usually perform the keyword process of nursing); that pills are involved, especially for sleeping; and that, since kidneys are particularly mentioned, some of the illness is serious. We might also notice – a kind of gap or ‘unsaid’ – that terms like recovery, well, better, health, are not listed as keywords or occurring collocates; thus we are not invited to use them in any narrativising reconstruction. In fact such positive terms are not on record (e.g., in the Collocations CD) as prominent collocates of words like nursing, pills and hospital: so any unfilled expectation of mention of ‘getting better’ must reflect cognitive schemas and scripts, rather than linguistic evidence. If the foregoing suggestions are reasonable, they also suggest that corpus and cognitive or background knowledge evidence can be profitably combined.

But what, indeed, is your or my script for ‘hospital’? Is it a place you end up in when you are seriously ill, or a place you leave when you recover from illness, or does it evoke some other internalised default ‘story’? This in turn raises a question for cognitive schema theory and its relation to corpus evidence: what better evidence is there, for any schema or script, than a weight of corpus evidence (collocational and colligational)? In fact what more reliable proof of scripts and schemata can there be than situated language use? And how is a cognitive schema revised or ‘refreshed’ (to use Cook’s term, who argues that literary discourse functions to refresh our cognitive schemata [Cook 1994])? Presumably there must always be inspectable evidence indicative of refreshment or change, of a communicative or behavioural kind, and such evidence will normally be most conveniently accessed in suitable corpora.

Several surprises, so often the best reward of corpus analysis, emerge: for example, my intuitions see no natural link between car and the items got and just. But in Section III, we find that the keyword just is also an occurring collocate of got, that both know and people are collocates of both these, and in particular that got in turn is a collocate of car. When so few items are harvested by the chosen technique of selection, this level of connectedness is remarkable.

Figure 2 also makes it more strikingly apparent that the reinforced sense of mutual relations among the keywords and their occurring collocates within each section does not seem to be matched by more evident or numerous links between sections. There are exceptions to this, such as the cross-linkages by means of mother, children, home and hospital between Sections I and II, and via the collocate people which figures in Sections II to IV. But these are comparatively rare. Some of this simply reflects the extent to which the different sections tell separate stories, with
little overlap between those told in Sections I, II, and III. On the other hand, there are continuities between the final three sections, and especially between II and IV, which are ‘Enid’s story’ and are focalised largely from her viewpoint. It may be that an over-emphasis on the distinctiveness of each section is the inevitable consequence when keyword analysis is conducted section by section, as here, by contrast with the ‘whole text’ calculation presented earlier. The ‘whole text’ analysis (14 lexical keywords), for example, identified car and mother as the top lexical keywords, and then listed said, water, riverbank and children. Satisfactorily, the section-by-section analysis (29 different keywords in all) registers only car and mother as key in two sections, and water in one, while said, riverbank and children disappear from view: they are key overall, but not in any of the sections taken singly.

It is arguable that a scenario such as that in Figure 2, which suppresses personal name and place name keywords (in addition to excluding grammatical items), leans thereby more towards theme than plot. A further expansion of the keyword base can be achieved by including in the ‘map’ not only the textually-occurring commonest collocates of keywords, but also any words occurring in the text that are plausibly regarded as semantically related to one of the keywords. On the other hand, this may introduce other frailties into the analysis. One might also attempt ‘soundings’ of smaller segments of text and their keywords – smaller than the (on average) 6,000 word sections studied here – in the hope of achieving a more delicate analysis. It may also be noted that in the figures above, nouns predominate over verbs, whereas narrative events are normally more centrally conveyed by verbs, so that the corpus analysis may need to be weighted in some way more towards identification of ‘key verbs’. According to Stubbs, who conducted a corpus analysis of Heart of Darkness, “Verbs are often a better candidate [sc. than nouns] for stylistically relevant words” (Stubbs 2005: 11; see also Yevseyev 2005 and Herman 2005). Even a focus on verbs, however, is only an intermediate step where the interest is in progression or narrativity: we need verbs (and nouns) embedded in postulated core propositions or narrative sentences of a kind that might ‘drive’ a narrative and guide our expectations concerning what is passing and to come. How a corpus stylistic narratology might best isolate these kinds of material will be taken up in Chapters 7 and 8. But before this, there are more things that need to be said about lexical patternings in “The Love of a Good Woman”, and about how corpus methods can throw light on some of these even while others continue to defy easy ‘automatic’ capture. The more one struggles to make corpus stylistics help us see things in the literary story, I believe, the more one learns about the complexity and subtlety of the finest stories’ structure and texture.
Lexical repetitions, partial repetitions, and degrees of similarity in wording, projected from the axis of selection to that of combination as Jakobson would have described it, have emerged as a preoccupation of this corpus stylistic exploration of narrativity. The present chapter again explores how the disproportionately recurrent words in “The Love of a Good Woman” might be a partial guide to its structure and progression. The analysis adopts a closer focus than the previous chapter, looking at the keywords in shorter (1,000-word) successive phases of the story. Interesting local effects come to the fore, such as the correlation of keyword personal pronouns with idiolect and free indirect speech in one, and exclusively in one, identifiably distinct phase of the story. The chapter’s chief argument is that repetition as narrativity guide is incomplete, and even has to fail for narrative (as distinct from several other discourse genres) to happen, because story ends must differ from story beginnings. This leads to a speculative discussion of whether, and how, a story opening or beginning, to which a middle and an end can be attached, is definable or identifiable by corpus stylistic means. Aristotle is often thought to have made a banal point in observing that stories have beginnings, middles, and ends. The final section of the chapter proposes a line of research that might be of value if we assume his observation was not banal but profound.

1. A more delicate keywords and plotlinks analysis of “The Love of a Good Woman”

In an attempt to get an even finer grained picture of the disproportionately-frequent items (lexical or grammatical) in the phases of the story, I divided the text into a total of 26 subsections, comparing their wordlists with the same 20th century fiction reference corpus wordlist as for earlier calculations. The story subdivisions were not arbitrary, but made with some attention to the evident structuring of the story itself. Thus the openings of each of the four authorial sections were preserved as section boundaries, and the short Preface was treated as a separate section. Then within the story’s four given sections, attention was paid to those
points where the author had marked an implicit sub-division by means of several blank lines, free of text; these always mark a shift of focalisation, or a major shift in time, and so it was reasonable to treat them as the boundaries of story sub-divisions. The entire story is 25,470 words long, and the sub-sections average a little under 1,000 words in length. But their actual lengths vary. Subsection 21, being the sequence of Enid’s journal entries, is exceptionally short: 145 words in total. These entries are presented belatedly and are so distinct from surrounding narration that I felt it best to assign them to a subsection of their own. The longest subsection is Subsection 11, which occurs at the beginning of Section II, and introduces the characters Enid and Jeanette Quinn: 2,631 words. All keywords calculations here are based on the log-likelihood rather than the chi-square formula: the latter yields impractically lengthy lists of keywords for each of these short sections.

Table 4. Positive keywords, subsection by subsection, ≥ 3 instances

| [Preface] | 1 | car, water, ditch, Willens, their |
| Section I 2 | 2 | They, or, bridge, Cece |
| Section I 3 | 3 | Or, boys, Salter, sirs, Bud, (home, they, fellows) |
| Section I 4 | 4 | Cece, eggs, lifter, Baptist, burner |
| Section I 5 | 5 | Bud, lumps, mirror, Bud’s, pie, potatoes, Doris, brother |
| Section I 6 | 6 | Jimmy’s, Cecé’s |
| Section I 7 | 7 | Jimmy, Cece, Cumberland, Jimmy’s, They |
| Section I 8 | 8 | Willens, forsythia, they |
| Section I 9 | 9 | Willens, phoned, Cece, Jimmy |
| Section II 11 | 11 | Enid, Quinn, Rupert, Mrs, Kidneys, Quinn’s, Ha. |
| Section II 12 | 12 | Enid, promise, nursing, mother, (or, Mother?) |
| Section II 13 | 13 | Enid, electricity, Mother, hospital |
| Section II 14 | 14 | Enid, Rupert, senior, cassava, crossword |
| Section II 15 | 15 | Enid, Quinn |
| Section II 16 | 16 | Enid, dreams |
| Section II 17 | 17 | Enid, Quinn, Mrs, Lois |
| Section III 18 | 18 | Rupert, Willens, Willens’s, car, cloth, Jutland, glug, got, she, up |
| Section IV 19 | 19 | [no keywords], |
| Section IV 20 | 20 | Enid, Lois, Quinn’s, punished, Sylvie, Quinn |
| Section IV 21 | 21 | July, Vy, Q, R, Eggnog, rain |
| Section IV 22 | 22 | Enid, green |
| Section IV 23 | 23 | Enid |
| Section IV 24 | 24 | would, will, Enid, swim, she |
| Section IV 25 | 25 | Enid, cone |
| Section IV 26 | 26 | Enid, Rupert, boots |
And it should be emphasised that by contrast with the keyword calculations reported in Chapter 5, which focussed on lexical items and strictly excluded proper names and grammatical items, this calculation of the keywords in each subsection imposed no exclusions.

In general shorter sections have fewer keywords, but this is no more than a tendency with exceptions. Thus the rather long Subsection 3 and the rather short Subsection 10 each have four keywords. Since the same item, typically a name, may be key in multiple sections, the total inventory of positive keywords, without lemmatisation, is 70 types, 136 tokens. *Enid* is key in thirteen subsections, *Cece, Quinn, Rupert* and *Willens* in four, or in 3. Numerous items are key in two subsections, including *Mother, and, Bud, car, I* and *Jimmy*.

The results for Subsections 11 and 17 include a useful reminder of how Keyword calculations can be analytic to the point of being misleading. In both those subsections it emerges that the items *Quinn* and *Mrs* are keywords, as is *Enid*. But what is not shown is that in nearly all occurrences *Mrs* and *Quinn* are adjacent items, in that order, referring to Jeanette. In a sense this is an extreme form of the tendency of keywords to be textual near neighbours, a tendency which merits further comment.

With WordSmith’s Plotlinks applied to the list of keywords for a section, it is possible to display the number of local co-occurrences of each keyword with any of the other keywords. The analyst determines how many words either side of a keyword (the node) the programme should search for a co-occurring keyword. I set the search-span at 5 words either side. Such co-occurring keywords are called links. But rather than examining all of them, it may be appropriate to focus on just those keywords which, for example, occur in a subsection at least three times, and which have at least three links with another keyword. Of particular interest is Subsection 18 (which also happens to be the entirety of Section IV of the story’s given structuring), Jeanette’s free indirectly reported confession of the sexual assaults and the killing. Just seven of the ten keywords that log likelihood has identified reach these thresholds (at least three uses and three keyword co-occurrence links). They are the following:

- *Rupert*, which links with *she* 7 times and with *got* 5 times.
- *Willens*, which links with *up* 4 times
- *car*, which links with *she* 4 times
- *cloth*, which links with *she* 7 times
- *got*, which links with *she* 9 and *Rupert* 5
- *up*, which links with *she* 8 and *Willens* 4
- *she*, which links with *got* 9 *up* 8, *cloth* 7, *Rupert* 7 and *car* 4
On this basis one might argue that *she* is the keyword that enters into the greatest number and variety of links with other keywords in this section. If all keywords are a kind of foregrounding by disproportionate occurrence, then those sentences in which several different keywords co-occur are especially foregrounded. Sentences in this 126-sentence subsection in which at least four different keywords from the above seven co-occur comprise just the following five (all keywords in bold):

*Rupert* banged his head **up** and down on the floor, *Rupert* banged the life out of him, and *she* jumped **up** so fast the chair went over and Mr. Willens's box where he kept his eye things **got** knocked over and all the things flew out of it.

*Rupert* didn't say anything at first, so *she* went into the kitchen and **got** some water and cleaned Mr. Willens **up** so he wouldn't dribble on anything.

They hoisted Mr. Willens **up**, *she* by the feet and *Rupert* by the head, and he weighed a ton.

After *she* cleaned **up** the floor, *she* could still see where *she* thought there was a stain, so *she* **got** the brown paint left over from when *Rupert* painted the steps and *she* painted over the whole floor.

So *she* **got** a cloth *Rupert*'s mother had embroidered with flower baskets and took it in there and *she* could still smell the smell.

Of course a longer list of sentences would be harvested if the threshold were set at just three different keywords in the same sentence. But the four-keyword sentences listed above are arguably strikingly relevant to the main story progression of this section, so that it could be claimed that a narrative's densely co-occurring keywords tend to be present in some of the sentences that are more central to the narrativity. But a good deal that might be judged important to the progression is not present in the sentences above, most notably any reference to Willens's assaults on Jeanette, or how Rupert disposed of the body. So again there is evidence of keywords making a significant but partial contribution to core narrativity.

2. Keyword personal pronouns and idiolect-signalling

As in the story analyses in previous chapters, no attempt has been made to prepare the text by removing any cohesion (cf. Hoey 1991). But whatever pro-forms do occur in the text are inevitably counted in Wordsmith's Wordlist and Keyword procedures, and where one or another pro-form is disproportionately frequent or infrequent in the analysed text, relative to the comparator corpus, then it will inevitably emerge as a keyword (as we have seen in the case of *she* in Subsection 18/Section III). So, in a sense, pro-forms and pronouns are not ignored.
In the Keyword analysis of “The Love of a Good Woman”, for instance, the 26 designated subsections reveal disproportionately high frequency of the pronoun *they* in Subsections 3, 4, 8 and 9, i.e. two early and two late segments of Section I of the story (Jutland), where closer scrutiny reveals that the pronoun usually refers to the three boys who are a kind of ‘collective protagonist’ in this section: they go to the river as a trio, jointly find the submerged car and Mr Willens’s body, together come back to town (where, Subsection 4 comments at length, to the townsfolk they are either boys or jocularly *sirs*), and almost reveal their news to Captain Tervitt. But they are not consistently a collective throughout Section I, and this explains the otherwise curious fact that *they* is prominent early and late in the Section, but not through Subsections 5 to 7 (or in the final Subsection 10). It is in these subsections that the narrative divagates to provide absorbing sketches of each of the boys’ home lives in turn, and what each of them encounters and does when they get home for lunch on the day of their newsworthy discovery. In terms that Munro herself has used, these are not so much side trips off the story’s royal road, but inspections of back rooms in the house of the story. Subsection 5 is entirely devoted to Cece Ferns and his family, Subsection 6 to Bud Salter, and Subsection 7 to Jimmy Box. The boys function collectively again in Subsections 8 and 9, which covers their walk through town and encounters with Mrs Willens (her *forsythia* is both a Section II and Subsection 8 keyword) and Captain Tervitt. But in the short Subsection 10 the boys again separate, and it is finally Bud who tells his mother about Mr Willens’s body in the river. So again, no prominence for the *they* pronoun.

In this discussion of keywords in the subsections I have not listed or attended to the very few negative (disproportionately infrequent) keywords that emerge, on the grounds that these are unlikely to be a reliable indirect pointer to the narrative drift. A perhaps inevitable limitation of negative keywords is that they draw our attention specifically to words that *do occur* at least once in a text, but do so markedly infrequently, by comparison with the reference corpus norm. But any contextually predictable lexis that is entirely absent from the analysed text will not and cannot feature as negative keywords. So negative keyword calculations cannot help identify lexical ‘gaps and silences’ in a text.

While negative lexical items may have slight informativeness, this may not be the case with grammatical items. Any negative grammatical keywords, like positive ones, may be much more discoursally revealing, since they are much less expectable than lexical ones. That is to say, where the reference corpus resembles the target text in genre and style, disproportional frequency of grammatical items should be rare indeed.

With that in mind, it is worth noting that the first-person nominative singular pronoun *I* is a negative keyword in Subsections 9 and 18, which amounts to saying, of this story without an overt narrator, that direct discourse is disproportionately
scarce in just those two subsections. In fact there is very little direct discourse, and only two uses of I in both these subsections, but the latter (Subsection 18) is much the more interesting case, being (as previously noted) the section comprising Jeanette Quinn’s shocking revelations in their entirety, expressed in free indirect speech (FIS) mode. As is widely recognised, FIS and FIT (free indirect thought) are remarkable hybrid forms, where the discourse seems to be entirely character-expressive, akin to direct speech or thought, in nearly all respects but tense- and pronoun-choice, which remain narrative-oriented, not character-oriented. Thus in a section which on independent grounds can be seen to report Jeanette’s direct and personal testimony, carrying her evaluations and not a neutral narrator’s, the marked absence of I may be treated as an automatically-derived indicator of extended free indirect discourse (FID: i.e., either free indirect speech or thought).

On the basis of such tellingly local trends, it may be reasonable to predict that an extended passage of FID has the consequence (in the orthodox narrative situation, of ‘third person, past tense’ narration), of causing a disproportionately heavy use of a particular third-person pronoun, even to the point of making it a positive keyword, as happens in Subsection 18 with she, and causing disproportionately light use of the first-person pronoun, even to the point of making it a negative keyword. In fact it may only be where a passage has one third person pronoun as positive keyword and I as a negative one, that FID may be indicated. Hence neither Subsection 9’s nor Subsection 24’s keywords indicate presence of FID: they have she as a positive keyword, but not in combination with I as a negative keyword. These first steps in the semi-automatic identification of FID text will be taken much further in Chapter 8, on the assumption that brief local passages of FID are usually sites of heightened narrative significance.

The fact that Section III (Subsection 18) is entirely Jeanette’s free indirect speech and thus very much in her idiolect also points to an explanation for the otherwise curious fact that in this section alone is up a keyword. This multi-functional item of course occurs elsewhere in the text of the story, but much less frequently (103 occurrences in 23,000 words) than in Section III (32 instances in 2,242 words; approximately every 70th word is up, or on average it occurs once every four sentences). The explanation is apparent as soon as a few examples are considered. Up is used in this section (and only here) in the colloquial and informal ways that are part of Jeanette’s idiolect:

Mr. Willens had been close up in front of her with the thing up to her eyes....her skirt got scrunched up.... The goo was not coming up in his mouth anymore.... Rupert had thought up the rest of what to do.....She said, A good thing, it’ll muddy up your track....she would have smashed it up, but how do you smash all those things in it?
Many of these uses of *up* are as a post-verbal particle, of slight independent semantic content over and above the sense carried by the lexical verb (*scrunch up, thought up, muddy up, smash up*); judging by the occasions in which Jeanette Quinn’s direct speech is reported in Section II, use of *up* as a post-verbal filler in this way is indeed part of her speech style. There is even one occasion there where Jeanette uses *scrunched up*: “Then Clive and Olive went in the truck and I could’ve scrunched up in their front seat but they never thought to ask me”.

3. The non-repetitive echo: long-distance patterning via associated lexis and analogy

There are aspects of lexical patterning, arguably important in the reader’s reception of this story, which a Keywords study cannot capture. They escape automatic identification even though they involve what might be called ‘para-repetitions’ (on an analogy with half- or para-rhymes): a phenomenon or image is described through one combination of lexical items at one point in the discourse, and seems evoked again in a comparable description at a quite separate stage through different (but related, synonymous, or associated) phraseology. Insofar as these are repetitions, they should be amenable to study as kinds of text-cohesive patterning; insofar as these ‘recurrences’ use differing lexical means, they elude identification via Keywords or similar search methods. But they are important: as expressions of elements that show sameness or continuity between an earlier and a later state along with evident change, they are part of the essential fabric of the narrative as narrative.

3.1 Rubbing and scratching surfaces

By way of examples of such intratextual links – later phrasings that can be seen both as repetitions of earlier ones and as new material – I will begin with two instances from the end of the text. These ‘echo’ just a few words and the scene or idea that these words project, which occurred in the story’s prefatory opening paragraphs. The relevant echoed sentences are those in the short final paragraph of the Preface, already much discussed in this chapter:

> Everything is black, but that is only paint. In some places where the optometrist’s hand must have rubbed most often, the paint has disappeared and you can see a patch of shiny silver metal.
Seventy-four pages later, Enid is posing questions to herself about the truthfulness of Jeanette’s shocking deathbed revelations. This final Section is titled ‘Lies’. Enid thinks about how lies and deceptions can develop in our tellings of events:

Look how elaborate dreams are, layer over layer in them, so that the part you can remember and put into words is just the bit you can scratch off the top.

Despite these instances being discoursally so many pages apart, one in the Preface set at a time many years after the 1951 depicted in Section IV, is there a connection between these two images and wordings, and some kind of allusion in the latter to the former? No simple text-search will highlight a link between these brief passages, and the lexis of the former (places, hand, rub, paint, disappear, patch, etc) is not repeated in the latter (layer, part, bit, scratch off, top), although here I am of course asserting a semantic relation. To be made explicit, that semantic or thematic relation needs to be worked out, by a truly active reader. The working-out involves something like the calculation of a veiled Gricean implicature, if not of a mathematical equation. It involves intermediate terms present in neither of the actual extracts – e.g. a word like surface (entailed by both places rubbed in extract 1, the part scratched in extract 2). In fact perhaps surface, face, layer, cover, top form a thematic network in the text. The poetic underspecification of the image in the second extract, where we might infer that Enid is alluding to skin or the scab on a partially healed wound, certainly adds to the uncertainty of the interpretive process. But the interpretive differences are as evident as the similarities: in the Preface image, rubbing has removed the ‘superficial’ black paint so that “you can see” the shiny silver visible; but in Enid’s remark about dreams, the deeper layers remain irrecoverable, and only the “bit you can scratch off the top” can be accessed and articulated.

3.2 Dark above, light below

The second example is a separate putative link between the lines from the Preface quoted above, and the following paragraph which occurs very close to the end of the story:

When she and Rupert went underneath the roof of summer leaves it was dusk, it was almost night. You had to watch that you didn’t trip over roots that swelled up out of the path, or hit your head on the dangling, surprisingly tough-stemmed vines. Then a flash of water came through the black branches. The lit-up water near the opposite bank of the river, the trees over there still decked out in light. On this side – they were going down the bank now, through the willow – the water was tea-colored but clear.
What I am particularly noticing comes in the first four lines, where in the surrounding dark “a flash of water came through the black branches” and lit up the water. In my view this is an intratextual echo of the earlier image of the black material with the shiny silver showing through: enough fellow readers have agreed that there is some such oblique connection to prompt me to claim that seeing such a link is a process of construal, or plausible interpretation, and not a merely imagined connection. Thus it is an element in the story’s texture, present now and for all time. Questions about how ‘fully consciously’ Munro planned and intended such a textural connection are unhelpful: in construing that link I and other readers have already given our answer to that question, as best we can. She has written what she has written, and it was not – we can assume – automatic writing. But the association is not starkly obvious, and there lies part of its strength. Its interest to reader and analyst is almost in proportion to the extent that it eludes standard corpus-analytic ‘capture’. There is no lexical common ground, for instance, between the cited passage and the opening paragraphs’ image, other than in the word black.

3.3 Sorrowful plummeting

A third example I will note more briefly concerns the images of sinking and drowning (or at least, of watery burial), both the reported one of Mr Willens, whose body inside his car was allegedly pushed into the river, and the fate that Enid imagines, near the story’s close, might befall her at Rupert’s hands. In dialogue with these, especially the former (since it occurs soon after it) are the following few lines, which occur towards the end of Section I, as the three boys walk past the Duke of Cumberland public house. We are told that this is where Cece Ferns’s father goes on payday, to get violently drunk. The text continues:

The word “Cumberland” always fell across his [Cece’s] mind heavily. From the days when he hadn’t even known what it meant, he got a sense of sorrowful plummeting. A weight hitting dark water, far down.

This is Cece’s reaction to his father’s drinking in the Duke of Cumberland; but it is also powerfully descriptive of what may have happened to Mr Willens’s body, even though it occurs at a point in the narrative where Willens is not remotely the focus of attention. Note, too, that at this stage in the story neither Cece nor the reader has grounds for assuming that Mr Willens’s death is other than a tragic accident or, possibly, suicide. Whereas if ‘weight hitting dark water’ is treated as interpretive of Willens’s fate (and not merely as expressive of Cece’s oppressive and melancholic thoughts about his abusive father), then it proleptically hints at more information than the reader as yet certainly knows, namely that Mr Willens was a dead or unconscious ‘weight’ when he sank into the water. Again, this is an intratextual
linkage that contributes story texture and the reader’s sense of theme, of the story’s priorities and even of possible progression (i.e., that there may be further ‘sorrowful plummetings,’ more weight hitting dark water, in the narrative ahead), but it is of a kind that eludes standard corpus-analytic tools.

3.4 Bashing, banging and braining

A somewhat contrasting fourth and final example concerns references to suffering a serious if not fatal head injury. These would hardly be interesting but for the fact that, assuming we believe Jeanette’s Section III confession, Mr Willens dies because Rupert banged his head up and down on the floor, Rupert banged the life out of him.

This seems echoic of some verbs and images that will hardly have delayed us in Section I, when we are told there of how in places the Peregrine River is life-threateningly engorged with ice-cold runoff:

It was barely back within its banks after the yearly flood that turned the flats into a lake and tore out the young trees and bashed any boat or hut within its reach.... if you fell into it, it would freeze your blood and fling you out into the lake, if it didn’t brain you against the buttresses first.

And then at the end of Section III, in the free indirect reporting of Jeanette’s confession, there is a graphic final description of Willens’s assaults on her, and shortly thereafter a mocking report of the authorities’ faulty interpretation of his death:

Right on the bare floor to knock her up and down and try to bash her into pieces... They said his head got bunged up knocking against the steering wheel.... What a laugh.

It is arguable, too, that the image or motif is echoed again at the very end of the story, in the passage already quoted in reference to Example 2, where the Enid-focalised narration comments:

You had to watch that you didn’t trip over roots that swelled up out of the path, or hit your head on the dangling, surprisingly tough-stemmed vines.

If this is an echo of the earlier dangers of head injuries, it is perhaps fitting that this late example is so attenuated, such a milder threat, and is represented as self-inflicted (“watch that you didn’t … hit your head”) rather than a sudden attack from an outside force.

All four of the above examples of suggested para-repetition, of varying importance to narrative structuring, escape mechanical capture by standard repetition-based searches. Is there some means by which they can be captured, drawing on
collocation-sensitive software and similar means? The prospects do not look very promising. For example, in response to the final postulated sequence of bashings and bangings, one might try to draw on a semantic tagset such as the UCREL tagging provided by Wmatrix (Rayson 2007), in the hope of finding all and only the automatically-identified subset of the text’s lexis that concerns the banging and bashing of heads. Even if such a subset emerged, it would be no more than a provisional first step in an unclear procedure. For instance, one might hope to find that this subset was disproportionately richly represented in the text, relative to other comparable subsets; but what then? One of UCREL’s semantic tagging categories, A1.1.2, covers terms that denote ‘damaging and destroying’, and 36 items from the story are automatically assigned to that category; unfortunately only one of the bashing/banging items listed earlier is so classified (the *bunged up* of They said his head got bunged up…). Proceeding in the other direction, one can see how Wmatrix semantically tags words like *bang* and *bash*; the former is classed as A1.1.1 (General and Abstract Terms: General actions, making etc.), the latter as E3- (Emotional Processes: Calm/Violent/Angry), so the linkage in their contextualised use in the story is not reflected in the groupings assigned by the semantic tagger. Similarly, the *rub* and *scratch* of the first pair of linked examples are in different categories in the semantic tagset: *rub* is grouped in with A1.1.1. general actions (like *bang*), *scratch* is B2- (health and disease). So it is difficult even to make a beginning on the predictive grouping of intuitively-associatable items.

There are methodical problems not only with identification of grouped instances by means other than close reading, but also with circumscribing each set of putatively-related instances, too. It is not easy to see a principled basis on which to include or exclude borderline examples. Thus in Section I when Bud is at home and a quarrel with his sisters erupts, the text runs “Bud ought to get his face smashed,” *said Doris*. Should this face-smashing be grouped with the other bashings and bangings noted above? And in the Jeanette-focalised Section III, soon after the previously-quoted description of Rupert banging Mr Willens’s head up and down, we are also told that *Rupert just walloped him*; and soon thereafter we are told how, presumably in his distress and frustration at what he has done, Rupert *started pounding on his knees* and Rupert *kept banging his big flat hands*. Should some or all of these, too, be included in the para-repetitive grouping?

The manner of proceeding adopted above, using instances noted via intuition in an attempt to develop a procedure for identifying an underlying link, may be regarded as far too *ad hoc* and contingent. Perhaps an automated search for widely-separated synonymous or near synonymous wordings would be preferable. But in a 25,000 word narrative, there is likely to be an enormous number of them, many of relatively slight prospective or thematic significance; they will rarely be instrumental in the reader’s experience of the narrative’s progression. And there is
no robust means of automatic identification and listing (let alone mapping) of all a text's lexical synonyms, to say nothing of near synonymy at the level of the phrase. We have also seen that identifying those words and phrases in a story's opening paragraphs that are likely to be crucial to progression and prospection is far from simply a matter of statistical extraction of the lexical keywords, whether those examined are for the story taken as a whole, or for author-designated section by section, or for a series of story subsections. From the example of disk onwards (the only fully lexical keyword in the story Preface), we have seen that not all lexical keywords are indicative of narrative progression, and that several other elements seem to be equally important indices.

4. Para-repetitive narrative bonding between story opening and remainder

The difficulty persists, then, of distinguishing and selecting the prospection-relevant ‘wheat’ from ‘the chaff’ – the majority of lexical resemblances which few would ever claim to be echoically significant to narrativity. A first (defeasible) assumption should be that the ‘seed’ of salient foregroundable echoic relations will be sown at the story opening, this position of most crucial importance, where what is to be told must ‘get started’ in the most effective way that the writer can devise. Words and phrases that echo ones used at the story opening also, logically, have the greatest opportunity of being long-distance ones, spanning the whole story. How far a story opening extends as a site for the generation of crucial terms in lexico-phrasal echoism is itself a difficult question, likely to require a different notional answer in every case. But for analytical purposes an initial norm of the first five percent of the entire story could be postulated – thus, in the case of “The Love of a Good Woman”, this would be the first 1,250 words of the story, comprising the Preface and the opening pages of Section I. That notional cut-off point does not accord with an obvious phase-boundary in the story, but a slightly larger segment does: the basic narration of the boys’ hike to the river for an icy dip, and their discovery of Willens's car with Willens inside it, dead, his hand poking through the opened top, comes to a close with the boys’ repeated amazed “Son of a gun”, 1,530 words into the story. This point marks the first authorial sub-division (marked by blank lines) in Section I; so this notional opening phase comprises the Preface and the first following sub-section, and is approximately 6% of the whole story. In fact, if the opening: remainder proportionality matches that of a paragraph's topic sentence and its remainder, then the average proportionality, with obviously wide variation, might be closer to 1: 10, rising to 1: 16 as here in longer stories, and dropping to 1: 5 in shorter stories. But these are no more than broad trends. Yet another
congruence for story Opening and Remainder might be Theme and Rheme, used in the systemic linguistic analysis of sentences in their textual function.

The task now is rigorously to select just those words or phrases in this 1,530-word opening that might be judged at least potentially central to (i.e., productive in) the narrative progression. Those relating to the Preface were discussed and listed above and will be listed again below, and it is inevitable that they also partially condition a reader’s narrativity judgements about the remaining paragraphs of this notional ‘Opening’.

Proceeding now to a sifting of the opening paragraphs of Section 1 on a similar basis of first considering the initial main simple clause or coordinated clauses of each paragraph (ten in all), the following clauses are selected:

This place was called Jutland.
The three boys … believed … that the name came from the old wooden planks…
There was a road, or a track, coming back from the township road, but it had never been gravelled, and appeared on the maps only as a dotted line, a road allowance.
The car tracks to the water’s edge on that spring morning were easy to spot but were not taken notice of by these boys.
It was colder here upstream than on the river flats close to the town.
So they would jump into the water and feel the cold hit them like ice daggers.
The tracks that they didn’t notice came right through the ditch.
There was a pale-blue shine to the water that was not a reflection of sky.
The car had a panel in its roof.
“Son of a gun,” these boys said.

Again narrative interest is drawn to the textual mention of what are particularised entities, conceptualised as independent wholes rather than subordinate parts of wholes, especially but not exclusively where those entities are human. These preferences seem to govern selections of certain phrases as most salient. As a result the conclusion the boys reach early in the long penultimate paragraph contains the information that conditions the reader’s attention to other paragraphs:

It seemed that inside the car a man’s body – it had to be the body of Mr. Willens – had got into a peculiar position.

By the time they reach this revelatory sentence, even if they have attended only to the initial clauses in the preceding narrative paragraphs, readers will have noted that three boys are involved, that they plan to “jump into the water and feel the cold hit them like daggers”, and that they do not notice the car’s tracks going right through the ditch. But all of this is superseded, or pushed down as crucial to longer-term narrativity, when we are told that the boys see that Willens’s body is
awkwardly lodged in the car, that his face was (even in life) “grotesque” without being “seriously intimidating”, and that they themselves feel somewhat empowered (“grateful”) in being the ones to discover this “catastrophe”. Another item which arguably will be harvested from these opening lines is reference to Mr Willen's face, which is noticeably fully described although it is not actually observable in the boys' narrative present. The boys have every reason to assume his head and face is, like the rest of him, directly in front of them; but all they can certainly see is one hand and arm. Like so much in the story, the visible and apparent is – presumably but not certainly – just the surface or top layer of complex things.

In short, the phrases and images representing Mr Willen's dead body in his car in the river are the 'top' narrativity material derived from these paragraphs; but they will not be the only material. There will in addition at least be notice of the three boys, and possibly also the idea of them jumping into the water, as this is announced as the purpose of their visit to the river; on the other hand, the use of the irrealis modal would (“they would jump into the water”) together with schematic or background knowledge about how people normally act upon discovering a dead body in circumstances such as those depicted will have undercut expectations concerning the ‘icy-swim’ event-progression by the end of this opening. The storyline of Raymond Carver's “So much water, so close to home” indirectly confirms this. There, three men find a young woman's body in the water at the beginning of their fishing trip, but complete their vacation before alerting the authorities: their community, like the story's readers, know this to be inappropriate behaviour.

Other first clauses in paragraphs will be set aside (as in the Preface) for their assumed excessive particularity, or generality, for narrativity purposes: e.g., There was a road, or a track, coming back from the township road, but it had never been gravelled. There was a pale-blue shine… Existential constructions like these will be expected to relate to setting, and stative elements. Along the lines set out in Chapter 5, they will only attract sustained reader interest where they introduce an entity of relative particularity and complexity, of which there is further proximate mention. On the other hand, some elements become more noticeable by virtue of their treatment: thus there are several negative particles in these paragraph openings, including two that emphasise that the boys did not notice the car tracks. I will hypothesise that these are salient for local narrativity, but less so for longer term purposes. For example, how prospective, across the span of the narrative, is the report that there was “a pale-blue shine to the water that was not a reflection of sky”? In a crime or spy fiction narrative an abnormality-reporting negative of this sort could easily be a clue to concealed information the full significance of which is withheld until close to the end of the story. But in this story the narrativity of this negative description arguably is spent very soon after, when the reader learns that the blueness is that of the car. Even here, though, more textual patterning may
be latent, beyond this very short-span gap and completion. This is because the word *shine* used here is morphologically related to *shiny*, mentioned in the image of black surface paint and shiny metal below, in the Preface. There is arguably therefore a more explicit suggestion here of what was implicit in that previous use, namely that surfaces can both conceal and indicate different and surprising facts.

The words and phrases from the 1,530-word ‘Opening’ that arguably have the greatest promise of relevance to narrativity or progression are:

*Willens, drown(ed), River, red box, instruments, escape, catastrophe, anonymous donor, optometrist’s hand, black, shiny (paint?), three boys, car, (man’s) body, Mr. Willens, face, grotesque, (hand?), respect/gratitude*

Now the test would be to see how many of these, and their collocates, and items semantically associatable with them, figure in the later phases of the story, while bearing in mind such factors as the following two:

1. since narratives are centrally about change, we can predict that some of these central terms, images or predicates will undergo change and may even be displaced in the course of the narrative. A vivid example of this is instanced by the key term from the Preface, seemingly confirmed later in this Opening by what the boys find, *drowned*. Only much later will we find that Willens almost certainly was not drowned at all, so while the entailments that Willens has died suddenly and even violently persist, the descriptor *drowned* ceases to be used.

2. making a broad distinction between the existential and the evaluative, a narrativity-key item like *Willens* or *body* or *box* can be expected to persist and recur in ways that more evaluative and descriptive terms, like *anonymous, catastrophe, grotesque*, and *gratitude* will not, in the same terms, recur. *Anonymous* is a cue for subsequent impressions of mystery, cover-up, not naming names, remaining silent as Enid finally seems to do, etc.; but we cannot reasonably expect *anonymous* or its near synonyms to be repeated.

What is the evidence that the lexical items and names identified above as the highest-prospectivity material in the notional 1,500-word Opening are indeed carried over and echoed, paraphrased, answered, etc., in the later sections of the story? The evidence is, as in previous strategies, mixed. Items like *Willens, drowned, river, body, paint, red, box, car, catastrophe* are ‘cashed in’ in later sections, as is reflected in the fact that several of them are identified as keywords, or occurring keyword collocates, in later sections; a few seem to have only a shorter-span and lesser subsequent resonance or quasi-repetitive ‘answer’. And a number of narrative strands and their indicative key terms only emerge much later, with little or no announcement in the Opening – most noticeably Jeanette Quinn’s illness, her
alleged sexual ordeal, and Rupert’s killing of Willens. More than these, there is Enid’s witnessing, her reactions, and the final dilemma she strives to resolve, the core of the story from the beginning of Section II, but lexically foreshadowed in only the most attenuated ways in the entirety of the Preface and Section I. In light of all these factors, and the sense that other aspects of narrativity besides keywords and the initial clauses in narrative paragraphs must be factored into the cueing of narrativity, the study of narrativity signalling via semantically-associatable repetition will be pursued no further here. This is not so much because it could not be continued, but because this chapter’s findings suggest that an exclusive focus on lexical frequency and repetition is too narrow an approach, and disproportional. The foregoing extensive exploration of repetition and keywords has helped to make apparent that a multi-factor modelling of narrative progression is needed, which embraces a variety of features or parameters that it is reasonable to see as contributory to narrativity. I outline such a model in the next two chapters.
This chapter briefly describes the main textual resources in the modern short story which, I propose, are particularly instrumental in guiding the reader’s expectations about narrative continuation. From the point of view of the reader I call this *narrative expectation*, whereas looked at as the elements of text design that cause expectation it is often termed *prospection*. While prospection and expectation are not quite the same thing, they concern a single set of phenomena: the text in its assumed or projected context, and the reader. Whatever the text in context prospects is (and must be) the basis and source of whatever the reader expects. The part played by the language of the text in the generating of expectations, and the part played by contextualisations of that text (the reader’s genre knowledge, background knowledge, etc.) are always hard to separate absolutely. This is because they do not start out as radically separate; rather, context-triggered expectations depend on textual hints and suggestions. A text must contain some words, some language, before it can give rise to contextual inferential interpretation by the reader. If, taking the contrary view, context preceded and determined text, the *raison d’être* of stylistics and linguistics would be radically undermined. Against this, an analysis such as the present one seeks to take the content and signalling of the narrative text as seriously as possible (while recognising that it is incomplete, i.e. that context and background also matter). Privileging the text and what it prospects in this way, we in effect declare ‘no expectation without prospection’.

For ease of presentation, discussion of the array of proposed narrativity ‘drivers’ is divided here into discussion of the more core parameters in this chapter, and commentary on the more embedded or delicate or subtler factors in Chapter 8. But the core/embedded distinction is intentionally not a standard pairing of terms (cf. core/periphery; independent/embedded): the eight parameters in these chapters are all one set, differing in degrees from each other rather than absolutely along one crucial dimension. Each of the features has been included on the assumption that it is central to the achievement of modern literary narrativity and prospection.

I also propose that all these features are normally attended to in *their punctuation-bounded sentential envelope*. That is to say, wherever a textual feature has been singled-out and deemed foregrounded for narrativity, it is not that feature
alone, but the whole sentence in which it operates that is here treated as the site of foregrounding. This is clearly a highly contentious axiomatic claim about the ‘prosodic scope’ or reach of particular foregrounded textual elements, and cannot be fully defended here (cf. Sinclair 2004: 194). But it has to be evaluated alongside alternative proposals, which might be much less satisfactory: for instance, that the reader somehow especially registers and attends to named characters, or narrative tense verbs, without any regard to their particular immediate co-texts. A text sentence is a typographically ‘complete’ and comparatively semantically robust (although hardly ‘free-standing’) segment of information or representation. In the vast majority of kinds of formal writing, the sentence is an authorially-determined ‘phasing’ of the message, to which readers in their text-processing (even their breathing, their points of pausing) must orient. It is a psychologically real segmentation in the reading of written narrative text. Only in exceptional cases, such as the later fiction of Beckett and Joyce, is the sentence as aid to phased processing partly or wholly denied to the reader.

Before proceeding, I want to make a special acknowledgement. In writing this chapter, the further I went in attempting to specify the matrix of putative core textual resources for creating and guiding prospection, the more I felt the influence, albeit at different points and in relation to different factors, of the work of Longacre, Labov, and Halliday re-surfacing; this chapter probably draws on their ideas in more ways than I could easily articulate.

The next sections embark on the discussion of key narrativity features, beginning with perhaps the most obvious ones: the naming of main characters and the use of finite narrative-tense dynamic verbs (action verbs). Each presents practical difficulties, of under- or over-reporting, for any semi-automatic identification procedure. For example, a story might introduce a main character by her proper name at the outset, but thereafter usually refer to her by pronouns (she, her, hers, they..) even though other characters also are denoted by those pronouns. How can an automatic procedure accurately count how many times that main character has been referred to in the story? I discuss my methods of addressing such difficulties.

1. Sentences featuring named main characters

1.1 High frequency and keyword character referencing

This parameter might appear an obvious and straightforward one; but as the discussion of the top keyword in Chapter 4 implied, there may be controversy over whether the focus should be confined to one keyword character’s sentences, or whether it should extend to several frequent keyword character names. By
‘individuals’ one can also include non-human animates (e.g., the fish in Hemingway’s “Old Man and the Sea”), and inanimate entities anthropomorphised or not – nations, great houses, boats, etc. Typically, however, the individuals are recognisably human, and their given name and relevant co-referring pronominal forms (Lenehan, he, him) will come high in a list of keywords for the story in which they appear. When this is the case, we have statistical confirmation of the prominence, in the given text, of ‘evokings’ of that character. Whether that prominence reflects directly on prospection or narrativity is harder to demonstrate, but also harder to doubt, because to do so is to postulate a narrative text in which an identified individual is repeatedly mentioned but remains peripheral to how we expect the story to proceed. That a repeatedly-mentioned individual may turn out to be peripheral is a possibility, but significantly, a remote one, a ‘red-herring’ reversal of expectation, something that may be exploited in some kinds of detective fiction, for example. But it is Griceanly reasonable for the reader to expect the often-named to be a centrally relevant story participant.

More narrowly, then, the focus should in principle fall on individuals denoted by long cohesive chains of co-reference (i.e., those persisting across most of the text). But which individualised characters in a literary narrative will turn out to be the ones repeatedly evoked via long cohesive chains is not something that readers usually know in advance: they must first read at least some way through the narrative (names in story titles are one obvious exception to this). They must simply take a punt on certain individuals, however they are named on early occasions (by proper name, pronoun, or definite description). Readers have to make a sensible guess (but a textually-guided one) that The man out of the last house mentioned early in Joyce’s “Eveline” will not figure largely in what follows, and the continuing absence of a second mention of him, let alone a cohesive chain of mentions, confirms this more and more firmly. Similarly “The Dead” starts with Lily, the caretaker’s daughter, but it is soon clear, through paucity of mentions, that she is a minor character in the story. It may be significant that, being something that the reader must hazard rather than be sure of in the course of a first reading, identification of keyword main characters is at first less secure and more provisional than reader-identification of some of the other features to be discussed in this chapter, such as the first sentence of each narrative paragraph, or the set of cues of free indirect thought proposed in the next chapter. But identifications of the second prospection resource to be discussed, narrative tense finite lexical verbs of action in clauses where a main character is a participant, obviously are contingent upon the choices made within the present parameter. Therefore those identifications, too, will be provisional at first. Before moving on to discuss such verbs, an excursus is needed to explain how I propose to deal, corpus-analytically, with the difficulty of tracking cohesive chains, through a text, of multiple references to a main
character by means of the occasional proper name and frequent pronominalisation. Extant corpus-analytic software has limited capability at identifying all and only those anaphors that co-refer with a particular textualised antecedent.

1.2 Modelling cohesive chains in long texts by sampling

How can a corpus analysis which is predominantly of lexical patterns take account of the sequences of co-referential pronouns in a narrative text? Such pronouns are not lexical themselves. Nevertheless they clearly contribute to referential patterns in harness with names and definite descriptions, usually occurring more frequently than the latter two. Elsewhere (Toolan forthcoming) I consider the strategy of preparing the text for analysis by removing ‘unwanted cohesion’ and ‘restoring’ full expressions (as in Hoey 1991), and argue that it is highly problematic in theory and in practice, besides running counter to the spirit of corpus linguistics, which is committed to using samples of real language as its evidential base. At the same time, there is no justification for simply ignoring all the anaphora found in a text: some taking account of it, and of how pronouns link by full or partial co-reference with full forms, is desirable. As for identifying and counting fully-ellipted forms, this is much more difficult; but these are far less frequent. In the course of the narrative passages of a story, as distinct from the direct speech, ellipses of character-denoting phrases mostly function only intra-clausally or intra-sententially, i.e. in highly localised links rather than in chains. These will not be considered further here.

With a long text (a long story or novel), some kinds of rough quantitative sampling of how pronouns are integrated with names may be desirable. For example, in a story such as Mansfield’s “Bliss”, with its main female character named Bertha Young, a Parts of Speech tagger (or a simple concordancer) can be used to find the total number of she, her, and hers pronouns. We can then use a sampling technique such as checking a random 10% (or ten items, whichever number is larger) of each of these forms for their apparent referent, in order to hazard a calculation of the proportion of each of these forms in the story as a whole that actually refer to Bertha Young (since a she or her in “Bliss” could alternatively be used to refer to various other females: Pearl Fulton, Mary the maid, Nanny, the baby, etc.). Below I show my working:

<table>
<thead>
<tr>
<th>Table 5. Trends in pronoun use in “Bliss”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total instances of</td>
</tr>
<tr>
<td>3rd Pers. Sg. Fem. pronouns</td>
</tr>
<tr>
<td>130</td>
</tr>
<tr>
<td>Proportion of 10% sample denoting Bertha</td>
</tr>
<tr>
<td>Number in full text likely to denote Bertha</td>
</tr>
<tr>
<td>Total instances of Bertha’s</td>
</tr>
</tbody>
</table>
WordSmith calculates “Bliss” to be 4,800 words long, in 376 sentences, averaging 12.76 words per sentence. Considering just the textual means of referring to Bertha by her name or by personal pronoun, we may hazard that there are approximately 185 (140 + 45) items that do so, i.e., on average roughly one every other sentence. This aggregation does not display the distribution, sentence by sentence, of these estimated 185 referring items, and there is the possibility that the distribution sharply fluctuates (e.g. She might almost invariably denote Bertha in some passages, but refer to various other female characters in one or several other passages). But here, for example, are the opening lines of the story, with paragraphs numbered, and with all cohesive links that co-refer to Bertha Young highlighted in bold:

(1) Although Bertha Young was thirty she still had moments like this when she wanted to run instead of walk, to take dancing steps on and off the pavement, to bowl a hoop, to throw something up in the air and catch it again, or to stand still and laugh at—nothing—at nothing, simply.

(2) What can you do if you are thirty and, turning the corner of your own street, you are overcome, suddenly by a feeling of bliss—absolute bliss!—as though you'd suddenly swallowed a bright piece of that late afternoon sun and it burned in your bosom, sending out a little shower of sparks into every particle, into every finger and toe?...

(3) Oh, is there no way you can express it without being “drunk and disordered”? How idiotic civilisation is! Why be given a body if you have to keep it shut up in a case like a rare, rare fiddle?

(4) “No, that about the fiddle is not quite what I mean,” she thought, running up the steps and feeling in her bag for the key—she'd forgotten it, as usual—and rattling the letter-box. “It's not what I mean, because—Thank you, Mary”—she went into the hall. “Is nurse back?”

“Yes, M’m.”

“And has the fruit come?”

“Yes, M’m. Everything’s come.”

“Bring the fruit up to the dining-room, will you? I’ll arrange it before I go upstairs.”

(5) It was dusky in the dining-room and quite chilly. But all the same Bertha threw off her coat; she could not bear the tight clasp of it another moment, and the cold air fell on her arms.

(6) But in her bosom there was still that bright glowing place—that shower of little sparks coming from it. It was almost unbearable. She hardly dared to breathe for fear of fanning it higher, and yet she breathed deeply, deeply.
She hardly dared to look into the cold mirror—but she did look, and it gave her back a woman, radiant, with smiling, trembling lips, with big, dark eyes and an air of listening, waiting for something... divine to happen... that she knew must happen... infallibly.

Now this cohesive network of the most easily identifiable items co-referring to Bertha Young by no means includes all the items in this opening passage that denote her. The complexities of mapping the other items reflect the complexity of literary text. Thus the you forms in Paragraphs 2 and 3 (What can you do if you are thirty...?) are part of a somewhat rhetorical inclusion of everyone, but in the first instance they clearly refer to Bertha. Bertha is the you, and your bosom denotes Bertha’s bosom or, in direct discourse, my bosom. Indeed direct discourse emerges in the fourth paragraph, where the I pronoun similarly denotes Bertha (with the you now referring to Mary). All of this proficient readers know, rapidly assimilating these shifting forms (Bertha as now a she, now an I) into their constantly adjusting identification of further references to Bertha. But it is quite a processing feat, and one that a novice reader might struggle over.

This standard practice – especially in written stories – of switching between narrating a character’s acts and words and presenting them directly means that a complex, at least ‘two level’, tracking of cohesive chains is involved. There are various ways of modelling or representing these kinds of shift of ‘world’ or stance, but the important point here is to recognise the at-least dual cohesive chaining. On one hand, the story character named Bertha features in an extended sequence of narrative or extradiegetic references by personal name, third-person pronouns, and descriptive noun phrases. On the other hand, there is a distinct chain of intradiegetic namings, vocatives and deictic first- and second-person pronoun uses, a cohesive chaining within the story world’s discoursing, rather than in the outer narrating of that world. Here, Bertha is an I to herself, and Bertha or you to Harry, Pearl and the others. These two cohesive chains, the extradiegetic and the intradiegetic, are aligned or taken together, without being ‘merged; in the creation and reception of the story.

The comments above avoid describing Paragraphs 2 and 3 as direct discourse, because in the context of following sentences, which are explicitly framed as what Bertha said or thought, these early paragraphs are more masked. Paragraph 2 could conceivably be a question thrown out by the narrative to the reader, and not unequivocally sourced in Bertha. I do not wish to dwell here on the subtleties of free direct thought and other categories of discourse presentation, but simply to recognise the potential referential openness of pronouns like you, as used in Paragraph 2. In view of that referential openness, I propose paying most attention to what predominates: third-person pronominal co-reference.
Taking all the above complexities into consideration, my working solution in my analyses has been to rely heavily on the naming of characters where it is both frequent and keyword. That dual condition excludes low-frequency keyword character naming. As we have seen, character-naming that is both frequent and a keyword in the story is usually a proper name (Corley), although it can also be a pronoun (as in “Eveline”) or a definite noun phrase (the blind man in “Cathedral”). A corollary of my working solution is that, where a character is merely pronominalised in an ‘important-seeming’ sentence, and that pronominal form is not highlighted as a keyword (as is the norm), then that sentence will not by those means qualify as crucial to narrativity. Instead, I rely on the likelihood that the sentence will qualify for inclusion in a high-narrativity abridgement by virtue of one or more of its other features. In effect, I am abandoning any special focus on pronoun-heavy cohesive chains as a core resource in the creation of narrative progression – notwithstanding the importance of such chains for other kinds of textual continuity and cohesive harmony. On their own account, I hypothesise, personal pronouns (and pronoun chains) in literary narratives are not a signal of prominence or deserving of special attention with regard to progression or prospection; rather the reverse, if anything. There is considerable linguistic and psychological evidence to support this approach, of which pronoun-ellipsis and ‘pro-drop’ languages are a part. In he came, he saw, and he conquered, for example, the second and third pronouns are so peripheral to narrativity that we feel little difference, with regard to narrative progression, in the version that removes them: he came, saw, and conquered. On the other hand, taken colligationally with other lexico-grammatical features such as narrative-tense modal verbs, personal pronouns may be highly significant (see Section 1 of Chapter 8).

2. Narrative-tense finite verbs in character-depicting action clauses

Throughout, my assumption is that everything in a text has the potential to prospect, but some things are more conventionally designed to prospect than others. These include verbs of action, where these are the nucleus of clauses involving (as Actor or Goal/Patient, for example) characters who have emerged as probable main narrative participants by keyword prominence of naming. From quite early in a story, we usually find that at least one of the main characters will usually have been disproportionately dwelt upon, with repeated mentions of them having been made. Where those mentions are adjacent to narrative verbs of action, so that we are reading clauses in which changes of state are being effected or undergone by a main character, it is reasonable to treat those clauses as central to narrative progression.
2.1 Narrative-tense verbs with inquits excluded

In a past tense narrative, intuition and experience powerfully suggest that certain clauses with past tense action verbs are crucial to narrativity. But we also know that very many other clauses with past tense action verbs are not. For example these, from “Bliss”: *Mary brought in the fruit; Nurse sat at a low table; The baby looked up at her again*. Tagging software cannot exclude these, however: it will identify all past tense action verbs indiscriminately. In this section I explain and justify some simple ways of selecting from the long list of past tense action verbs that an automatic search identifies. My chief proposals are to exclude all ‘action’ verbs which are actually verbs of communication (such as the many *he saids* and *she saids* to be found in written narratives), and to exclude all verbs where a keyword-named main character does not co-occur very close by textually, usually as Subject or Object.

A first listing of all the lexical narrative-tense verbs of a story can be made using a program like Wmatrix, which has a category VVD for anything it recognises as a past tense lexical verb. Non-finite verbs, the VVI category in Wmatrix, are of some importance, but my assumption will be that they are less directly plot advancing. The VVD list of verbs in the story “Bliss” reveals that *said* is by far the commonest (36 uses), but many of these have the nurse or Norman or Eddie as sayer, and are intuitively less central, less contributory to narrative expectations than to other uses. This is most strongly felt in the little narratives told by the guests about their journeys to the dinner by train or taxi, and the simian characters they have encountered on the way. By contrast, anything said by Bertha or Pearl or Harry potentially is so (and their personal names are both frequent and key keywords, unlike those of Norman and Eddie).

Then again, among the reports of things said, we surely need to distinguish between

> "My little precious," said Bertha.... You're nice you're very nice!" said she.

and instances such as the following two (non-adjacent in the text):

> "Have you a garden?" said the cool, sleepy voice [of Pearl Fulton].
and with her eyelids Miss Fulton said: "Yes."

What might be the basis for judging the first of these as having next to no narrativity or prospective relevance? Everything demonstrated by discourse analysis of the variable realisations of speech acts suggests that the form of Bertha’s utterance is an unreliable guide to function, and that content and context must be considered. It is because we know that Bertha is addressing a baby, who cannot reply, and because her remarks are largely phatic, and because they are in context neither contrastive nor directly implicative of future steps in a narrative, that we assign low
narrativity to them; mainly the remarks are a further expression of Bertha’s blissfulness. Similarly the form alone of Miss Fulton’s “Have you a garden?” might prospect no more than a simple “Yes” from Bertha, thus contributing merely to the static description of the narrative setting by indirect (non-narratorial) means. But the reader already knows Bertha has a garden and, besides, in the post-dinner context of Miss Fulton’s utterance the reader interprets it as Bertha herself does, as a request to be shown the garden. Like any request, even if the addressee rejects or deflects it, this is richly narratively prospective: we expect Bertha to show Miss Fulton the garden, we may even expect them both to dwell on the pear tree, and Bertha may be expecting this episode to lead to the giving of a ‘sign’. As for the second example, Miss Fulton’s eyelids saying “Yes”, this is interactionally retrospective in an obvious sense, a confirming of a proposal raised as possible and negotiable (expressed in Harry’s preceding gestures to her, and his whispered “Tomorrow”). But like any agreeing to a proposal (as distinct from the ‘Yes’ that affirms a proposition: compare “But how dreadful!” she cried. “Yes, it really was,” said Eddie) it prospects future actions that will realise what has been proposed. Again, the reader’s judgement that here is an agreed-to proposal and not an affirmed proposition draws on rich contextual evidence and schematic knowledge (or what Werth 1999 calls ‘common ground’).

All the above considerations taken together lead me to treat all reporting or inquit VVD verbs (here, said, cried, thought, etc.) as in themselves of lesser significance (even though numerous), and only part of the foregrounded prospection text where they are sententially linked to direct discourse which is itself prospective. Thus the points made in the previous paragraph also relate to the prospective power of characters’ direct speech, and this quite separate resource is treated in full, as a distinct narrativity parameter, in Section 2 of Chapter 8.

By excluding inquit verbs and their complements, a small proportion – perhaps a fifth – of the 256 VVD verbs found in ‘Bliss’ (by Wmatrix) can be removed from consideration; this group includes said (35 instances), cried (4), called (3) whispered (3), murmured (2), and single uses of verbs such as reminded, warned, pleaded and stammered. In fact it is possible that the VVD list could be stripped automatically of any instances of a core inventory of listed reporting verbs. This still leaves 200 VVD verbs to be assessed for their salience to narrativity. But by means of the first postulated narrativity resource--sentences containing frequent character-naming keywords--many sentences will already have been selected where a VVD verb involving a named character occurs. Chiefly then the need is to identify those high-prospection VVD verbs used in sentences where no main character keyword name is used, perhaps because the individual happens to be pronominally denoted, or because the reported event or development does not
directly involve a main character. Such sentences may include reports of ‘acts of God’, coincidences, and partial or even total narrative surprises.

By way of preliminary indicative sampling, if the first 50 VVD verbs are taken in order of occurrence from “Bliss”, we find that approximately 27 are excludable (and are here marked EXCL) from further attention in this phase, on the grounds that they serve to introduce speech or thought, or lack reference (e.g. by name or pronoun) to a main (keyworded) character as Subject or as other inner argument. One further ‘instance’ is excluded: the adjectival phrase *shut up*, erroneously tagged in Wmatrix as a past tense verb. Here are those first 50 VVD verbs, with their immediate co-texts:

She wanted to run…….The sun burned in your bosom….It shut up in a case (not VVD)….She thought EXCL….She went into the hall….Bertha threw off her coat….The cold air fell on her arms …She hardly dared to breathe….Yet she breathed deeply….She hardly dared to look into the cold mirror….It gave her back ….She knew must happen…..Mary brought in the fruit EXCL….She stood away from the table….The table seemed to melt EXCL….She began to laugh….She seized her bag…(she) ran upstairs…Nurse sat at EXCL…[Baby] She looked up EXCL/ by interp …when she [baby] saw her mother EXCL/ by interp …Baby(she) began to jump EXCL/ by interp …said nurse, EXCL …that Bertha knew,…and that meant she had come…whispered Nanny EXCL…We went down to the park EXCL … I sat down EXCL …And took her out EXCL…Big dog came along EXCL…And put its head EXCL….And she clutched its ear EXCL…Tugged it EXCL…Bertha wanted to ask ….She stood watching….The baby looked up at her…. [baby] Stared EXCL?…Smiled so charmingly EXCL?….Said Nanny EXCL….Said she EXCL….Nanny handed her over EXCL…Nanny went out of the room EXCL…Said Bertha EXCL…[baby] She ate delightfully EXCL……she waved it away EXCL….Bertha turned round to the fire…Said she, EXCL…She loved Little B so much….She bent forward…. [toes] shone transparent EXCL…..

Except in the case of coordinated verb phrases (in which case the sentence will be selected by virtue of the proximity of character naming and verb in the first element: e.g. *She seized her bag and ran upstairs*), there seems to be a good chance that extraction of the verbs and sentences one does wish to retain may be performed semi-automatically:

Having used Wmatrix to identify all VVD verbs, retain just those VVD-verb-containing sentences which are neither direct speech nor a reporting construction (i.e., an *inquit* or verb of communication) and where, without intervening punctuation, a pronominal denoting of a character (for Bertha this will nearly always be she, her) co-occurs within two words to the left or four to the right of the VVD verb.

By this procedure, all those instances one wants to retain would be selected, including *The baby looked up at her*, where a main character is only indirect object
Unfortunately a very few other VVD-containing clauses would also be retained, including such sentences (referring to the baby) as she ate delightfully and she waved it [her food] away. Ideally, then, some means of excluding these instances which only involve a pronoun which is not co-referential with a main character (Bertha, Pearl, or Harry) is needed; but in fact no automatic or mechanical means looks likely to succeed here.

Co-occurrences of a VVD verb and any proper name can be ignored here on the assumption that relevant ones will have already been captured by the first narrativity criterion, described in Section 1 above, which privileges sentences featuring named main characters. Thus Bertha, Harry and Miss Fulton are character-naming frequent keywords, so that by my first criterion every narrative sentence containing their names is deemed crucial to prospection. We therefore do not need to strive to select, independently by the second criterion, a sentence such as Miss Fulton moved towards the hall and Bertha was following when Harry almost pushed past. This one would have been already selected — three times over, by virtue of its three keyword character names. One phrasing that will be missed is where a non-keyword-using description has been used in place of a name or pronoun, such as (modified from the text) when the baby saw her mother. Also the above procedure would not, among the verbs listed above, identify [that late afternoon sun] burned in your bosom, but then it is a moot point whether it should. Besides, we will find this entire sentence can be independently captured via the second step in the FIT-identifying procedure (see Chapter 8, Section 1).

The Wmatrix VVD listing does not include complex verb phrases where the past-tense marker has floated left to an auxiliary or operator: but she did look…she did not dare to (ask)...she had come. But the first of these is a rare emphasising construction, the second would be highlighted independently by virtue of its negative element (see Section 3 of Chapter 8), and the third is arguably less narratively dynamic, not on the ongoing narrative time line, in being perfective or distant past.

2.2 VVD density

Even with the various constraints and exclusions proposed above, the treatment of the nature of the subset of narrative tense action clauses that are crucial to progression is incomplete and needs further refinement. One tentative observation that may be offered at this early stage is that, among those clauses (or sentences) of particular interest should be those passages where multiple non-*inquit* VVDs arise
in comparatively close proximity, that is, with increased density. Consider, for example (bold added):

Bertha realised that she [Pearl] not only bored him [Harry]; he really disliked her. And she decided from the way Miss Fulton said: “No, thank you, I won’t smoke,” that she felt it, too, and was hurt.

And:

He tossed the coat away, put his hands on her shoulders and turned her violently to him. His lips said: “I adore you,” and Miss Fulton laid her moonbeam fingers on his cheeks and smiled her sleepy smile. Harry’s nostrils quivered; his lips curled back in a hideous grin while he whispered: “Tomorrow,” and with her eyelids Miss Fulton said: “Yes.”

In other words, sentences in the narrative containing two or more action verbs involving a prominent character are, we may hypothesise, more dynamic and therefore more crucial to progression than those containing just one.

Rises and falls in VVD density should be relatively easy to calculate, even without attention to sentence boundaries. If there are 4,800 words and 256 VVD items in “Bliss”, the average frequency is one VVD per 19 words. But in the first of the two passages just cited there are 5 VVDs in a span of 37 words (was, not being a lexical verb in the Wmatrix scheme, is not part of the VVD class), and in the second, 7 action verbs in 61 words; in both cases the VVD density (more than twice the story average) is woven in with direct speech and reporting verbs.

It is perhaps not sheer accident that I selected those two passages, for arguably they are the high points of the narrative – or the second one is revelatory high point, thrown into particularly dramatic relief by the first. In addition to being verb-dense, the passages palpably contradict each other (their prospections clash): in the first, Bertha judges that Harry dislikes Pearl, but within a few lines, in the second, we find he adores her. Such explicit oppositions (about a topic of such importance to the three protagonists involved) are an eloquent expression of change or changed state, the crucial requirement in narrative. The changed state is in Bertha of course – this is not a narrative in which Harry and Pearl become lovers – whom we are told first thought things were one way and now finds they are all a different way. So the thematic and narrative importance of these short passages, in particular, supports the idea that sentences with high action-verb density are foregrounded for narrativity.
2.3 Is independent capture of narrative-tense action verbs needed?

Interestingly, all but one of the sentences in the extracts just quoted would have already been selected by the first prospection-foregrounding procedure, described previously. This is because all of these sentences contain at least one of the keyword main character names: Bertha, Fulton, and Harry, the story’s top three keywords, each of them more than 0.4% of the text, with keyness scores over 100. All, that is, but for one exception:

He tossed the coat away, put his hands on her shoulders and turned her violently to him. (bold added)

In light of this, rather than independently searching for VVD verbs, one might argue that it will be sufficient to highlight those VVD verbs that occur in text that is highlighted by some other criterion (such as that of being a sentence containing a frequent character-naming keyword – Parameter 1 – or, to mention criteria to be discussed below, negation-carrying, or internal-thought expressing sentences). At least, the above looks to be a promising strategy, except for sentences like He tossed the coat away…. It is true that character-keyword sentence extraction will pick out the previous sentence,

And she saw... Harry with Miss Fulton’s coat in his arms and Miss Fulton with her back turned to him and her head bent.

And that sentence, together with the next one selected which begins His lips said: “I adore you”, certainly conveys the main point of what Bertha glimpses. But it would still be desirable to find some principled basis for including the intervening He tossed the coat away sentence, with its shocking confirmation of what seems unremarkable or at worst ambiguous in the previous sentence, where Harry only has the coat and not Miss Fulton “in his arms”. This is clearly a topic for further research; here I will not venture beyond suggesting that the second and third clauses of the target sentence are intuitively more important than the first, about tossing the coat away. If it is agreed that turned her violently to him is the most narratively significant part of the target sentence, its significance and a means of highlighting such sequences might be to treat as foregrounded those passages where a VVD verb (especially a material process one) is closely followed by a main-character-denoting name or pronoun, i.e. where the verb is likely to be transitive, and the following phrase is likely to be Goal or Medium. It should be possible, using a modified and simpler version of the extraction procedure given in 2.1 above, to automate a search for all VVD’s closely followed (within two or three words) by a third person possessive or object pronoun. Even then, determining whether any pronoun so identified refers to a main character or someone else will still require human interpretation.
3. The cueing power of first sentences of narrative paragraphs

A third resource in the cueing of prospection or plot-development, I propose, are the first sentences of narrative paragraphs. Paragraphing has long been recognised as an attention-focussing segmentation of the written text, and there is a steady interest in its stylistic effects (e.g., Emmott, Sanford and Morrow 2003, 2006; Hoey 2005: 129–148). Since a new paragraph is always normally entailed when a new story section occurs, the focus henceforth will be on narrative paragraphing and the re-orientation that may be performed by the paragraph’s opening sentence. But paragraphing is an articulation that may vary from author to author, and depending on genre and topic. Chafe (1994: 298–9) contrasts the short, one-sentence, paragraphing of tabloid journalism with pages-long paragraphs in The New Yorker, and talks about “fine- to coarse-grained paragraph marking”. Hoey (2005) finds, in informants’ paragraphing of a passage comparing Generals Grant and Lee, that some paragraph breaks seem almost required, while others are more a matter of personal preference. As for short story paragraphing, two stories of equal length might have few or many paragraphs, without it following that the story with high-frequency paragraphing also had a greater quantity of significant changes of setting or perspective, or developments in plot. Narrative-paragraph-initial sentences are thus a signal of narrativity, but a fallible one, to be used with caution.

Deciding just what blocks (and extents) of text to treat as genuine narrative paragraphs is also open to argument. In “Two Gallants” some long and many short paragraphs in the story are largely or entirely ones of direct speech, signalled by Joyce’s characteristic long em dash. The first instance of direct speech in the story, for example, is:

– Well!... That takes the biscuit!

In the analysis of “Two Gallants” into narrative paragraphs, I have de-selected all such direct speech paragraphs and have moved forward to the opening of the next narrative paragraph. I have also discounted very brief mid-dialogue ‘paragraphs’, where one or two sentences describing a character’s manner of speaking are supplied.

Here, with admittedly several difficult decisions as to inclusion or exclusion, is the sequence of first sentences of the narrative paragraphs in “Two Gallants”:

The grey warm evening of August had descended upon the city and a mild warm air, a memory of summer, circulated in the streets.
Two young men came down the hill of Rutland Square.
He became serious and silent when he had said this.
Corley’s stride acknowledged the compliment.
Lenehan offered his friend a cigarette.
His bright, small eyes searched his companion’s face for reassurance.
They walked along Nassau Street and then turned into Kildare Street.
The two young men walked up the street without speaking, the mournful music following them.
At the corner of Hume Street a young woman was standing.
Corley did not answer.
Lenehan observed them for a few minutes.
Lenehan walked as far as the Shelbourne Hotel where he halted and waited.
Now that he was alone his face looked older.
He walked listlessly round Stephen's Green and then down Grafton Street.
He was hungry for, except some biscuits which he had asked two grudging curates to bring him, he had eaten nothing since breakfast-time.
He spoke toughly in order to believe his air of gentility for his entry had been followed by a pause of talk.
He paid twopence halfpenny to the slatternly girl and went out of the shop to begin his wandering again.
He left his friends at a quarter to ten and went up George's Street.
His mind became active again.
Suddenly he saw them coming towards him.
They turned down Baggot Street and he followed them at once, taking the other footpath.
Lenehan hurried on in the same direction.
Corley turned his head to see who had called him, and then continued walking as before.
He came level with his friend and looked keenly in his face.
They had reached the corner of Ely Place.
Corley halted at the first lamp and stared grimly before him.

Like the top-keyword abridgement, this sequencing is to a degree coherent and informative, in ways that should not entirely surprise us. One of the things that is most noticeable about this sequence of sentences is that, after some awkward transitions between the early sentences (which may reflect more on which sentences I have identified as ‘paragraph-initial’ or not), the cohesive referencing is mainly remarkably clear and unproblematic. In other words, if these were the only sentences of the text that a reader attended to, they might coherently grasp the plot of the story (but not its theme or point). Thus Lenehan is mentioned by name in Lenehan walked as far as the Shelbourne Hotel, and the sequence of he pronouns that follow all refer to him. Only the antepenultimate sentence, He came level with his friend…remains referentially ambiguous. One thing this abridgement by paragraph-initial sentences completely obscures is the main plot ‘problem’ of the story – Lenehan wanting money secured by Corley from the woman – and the mild suspense this gives rise to: has Corley pulled it off, Lenehan wonders. But the paragraph initial sentences are a good guide to the ‘blocking’ of the action, a tracking
of the main movements and shifts in perspective (including the shift from the external to the mental: e.g., from *He left his friends at a quarter to ten* to *His mind became active again*). And since only two of these twenty-six sentences have an initial conjunct or disjunct, perhaps those markers, and their following sentences, are also especially foregrounded: *Now,* and *Suddenly.*

4. Narrativity carried by “fully lexical” frequent keywords and clusters

4.1 Frequent keywords

Multiple namings of main characters is the prototypical way of conveying their continuing relevance to the story, and finite narrative-tense action verbs are the core means of reporting the developing action. But another reflex of a story’s core narrativity is apparent where a text has frequently-occurring keywords other than particular grammatical items and particular names. That was the overarching thesis of two previous chapters. In practice, due to the brevity of the text, some short stories yield very few keywords that are both fully lexical and reasonably frequent (‘reasonable frequency’ is defined later in this section). When they are present they merit attention.

In the case of “Two Gallants” there are arguably just two frequent lexical keywords, *street* (26 instances) and *walked* (12). Other lexical items in the keyword list are quite low frequency and with low keyness scores (as noted in section 4 of Chapter 3): e.g., *tram, companion’s,* and *peas.* If we first ‘enrich’ the haul for *street* and *walked* respectively by noting occurrences of grammatically-related forms, i.e., by lemmatising, then we can add to these, respectively, instances of *streets* (3), and *walk* (3) and *walking* (6). The Cobuild collocates CD reports that *walk* is one of the top lexical collocates of *streets* (but not of *street*), and that *street* is a top lexical collocate of *walk, walking* and *walked.* And in “Two Gallants”, similarly, *street* is a top lexical collocate of *walked.*

Concordance lines for *S/street(s)* in “Two Gallants”

1 in a public-house in Dorset *Street.* Most people considered Lenehan id, I was going along Dame *Street* and I spotted a fine tart under
defy in a house in Baggot *Street.* I put my arm round her and
gaze after someone in the *street,* it was necessary for him to move
3 I saw her driving down Earl *Street* one night with two fellows with
4 They walked along Nassau *Street* and then turned into Kildare Street
5 nd then turned into Kildare *Street.* Not far from the porch of the club
6 young men walked up the *street* without speaking, the mournful music
7 ley. At the corner of Hume *Street* a young woman was standing.
Chapter 7. Prospection and expectation

Here? – Corner of Merrion Street. We’ll be coming back.

As he approached Hume Street corner he found the air heavily

and then down Grafton Street. Though his eyes took note

at ease in the dark quiet street, the sombre look of which suited

warily up and down the street, went into the shop quickly. He

again. He went into Capel Street and walked along towards the City

. Then he turned into Dame Street. At the corner of George’s Street

t. At the corner of George’s Street he met two friends of his

ur before in Westmoreland Street. At this Lenehan said that he

een Mac in Westmoreland Street asked was it true that Mac had

o ten and went up George’s Street. He turned to the left at the

and walked on into Grafton Street. The crowd of girls and young

ned and on his way up the street he heard many groups and couples

ed the corner of Merrion Street he took his stand in the shadow

ip. His eyes searched the street: there was no sign of them.

They turned down Baggot Street and he followed them at once,

e left and went up the side street. His features were composed in stern

of summer, circulated in the streets. The streets, shuttered for the repose

circulated in the streets. The streets, shuttered for the repose of Sunday

He had walked the streets long enough with friends and with girls.

Some local patterns emerge from these concordance lines: typically, street is recur-
rently immediately preceded by a name (Dorset, Dame) or the definite article, and
this in turn is preceded by one of a small set of prepositions (in, along, down, into,
up), and this in turn is preceded usually by a pronominal subject (mainly he) and
a past tense finite verb of movement (such as turned, walked, went). So there is a
quite prominently recurring pattern in this story (see Figure 3 below).

Like any collocational pattern or lexical item (Sinclair 2004), this tends to have
weaker semantic associations and colligations at its outer limits. Here, the increas-
ingly weak links are to the left of the sequence only, since street is a decisive right-
ward boundary to the pattern. For example, in none of its 26 occurrences here is it
postmodified, as in on the street where she lived.

<table>
<thead>
<tr>
<th>He</th>
<th>walked</th>
<th>into</th>
<th>the</th>
<th>street</th>
</tr>
</thead>
<tbody>
<tr>
<td>They</td>
<td>went</td>
<td>down</td>
<td>Dorset</td>
<td>Street</td>
</tr>
<tr>
<td></td>
<td>turned</td>
<td>along</td>
<td>Grafton</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Collocational trends before Street in “Two gallants”
Here now are all the instances of walk/walked/walking:

Concordance lines for walk/ed/in in “Two Gallants”, in story order

3. To a close. The other, who walked on the verge of the path and
13. you know. So we went for a walk round by the canal and she told me s
19. is father's frame and gut. He walked with his hands by his sides
22. word. He was often to be seen walking with policemen in plain clothes
24. arette. As the two young men walked on through the crowd Corley
40. ion; that's what she is. They walked along Nassau Street and then
43. and full. The two young men walked up the street without speaking
51. for a few minutes. Then he walked rapidly along beside the chains
57. position of his hat. Lenehan walked as far as the Shelbourne Hotel
58. side of Merrion Square. As he walked on slowly, timing his pace to
61. after each group of notes. He walked listlessly round Stephen's Green
64. passing them but to keep on walking. He turned to the left when he
72. n he beheld the pair of lovers walking along some dark road; he heard
75. inner to sit down to. He had walked the streets long enough with friends
77. He went into Capel Street and walked along towards the City Hall.
78. that he could rest from all his walking. His friends asked him had he
82. left at the City Markets and walked on into Grafton Street. The crowd
90. read the result in their walk. They were walking quickly,
90. result in their walk. They were walking quickly, the young woman taking
94. sed on her and Corley began to walk swiftly towards Stephen's Green.
96. called him, and then continued walking as before. Lenehan ran after him

The 21 instances of walk/ed/ing show that besides walked featuring quite often in the lexical pattern that has street at its nucleus, all three forms of walk here enter into recurrent sequences where, following the verb, there is either a post-verbal low-semantic-value item such as on or along, or a true adverb (rapidly, slowly, listlessly), or both. What are noticeably absent are reports of walking to a stated destination, purposively. This is most apparent when Lenehan is Subject and Actor of the walking: we find he walked on into Grafton Street (with no implication that somewhere in Grafton St is his goal) but never he walked to Grafton Street.

All the reasonably frequent fully lexical keywords in a short story text deserve attention, although as noted in some stories there may be very few items that fit this description. Where a lexical item is identified as key but actually occurs only a handful of times, or, more precisely, where the type comprises less than 0.1 % of the text's words, it is hard to argue for its significance to narrative progression. Therefore it seems desirable to impose a threshold below which even an item identified as a keyword in the text will be deemed insufficiently prominent. For a lexical keyword to be sufficiently frequent to be taken account of in the model of narrative
prospection, I suggest it must occur at least 5 times in the text, or comprise more than 0.1 % of the tokens of the text (a frequency of one word in a thousand).

None of the foregoing intends to deny that the single-occurrence word may be profoundly significant in a narrative. Equally, lexis from the language’s strata of moderately frequent items, intermediate between the commonest and the rarest, may be particularly revealing of the stylistic stamp of particular authors: see, e.g., Moon’s study of Conrad (Moon 2007). Particular uniquely-used words (termed *hapax legomena* in corpus studies) may be powerfully contributory to theme or outcome, and in a different study might take centre stage. This remains the case even though distinguishing a minority of especially significant single-use words from among the majority which are not presents its own difficulties: in most texts, more than half the text comprises words occurring only once. If such special single-use words contribute to narrativity at all, it is assumed here that at most they will contribute to peaks and reversals and endings in narratives, rather than to the main, broad prospected progression. So they may ‘cap’ or figure in the completion of a narrative surprise, for example, but are assumed not to contribute to the conditions in which a surprise emerges.

Finally it has to be conceded that short stories are often such small samples that they may fail to yield an informative set of keywords at all; and particular care also needs to be taken to ensure that prominences are not concealed by reliance on a lexically-skewed reference corpus. A difficulty of the latter kind arose when using my half-million word reference corpus for comparative identification of keywords in Mansfield’s “Bliss”. Against that corpus of modern narrative fiction, the following keywords emerged: *Bertha, Fulton, Harry, Norman, Nanny, taxi, pear, Knight, Warren, she, bliss, M’m, Mug, Eddie, Baby, she’d* and (negatively) *he*. The only ones that are not names or name-like are *taxi* (5, plus 1 *taxis* and 1 *taxi-man*, but all within the direct speech), *pear* (which always co-occurs with –tree; there are 8 instances,.17% of the text), and *bliss* (6 instances,.13% of the text; plus 1 *blissful*, all in the narrative text). Neither *bosom* (3) nor *cat* (3) is key; but more surprising to me was that *silver* did not emerge as key either, even though Wordlist shows that it occurs as frequently in the story as *pear* (8 occurrences). Applying Wordlist to the fiction reference corpus reveals that *silver* is quite common therein, with a frequency of 0.02%, or 2,000 per million; although *silver* is eight times more frequent that this in “Bliss”, it was not sufficiently more frequent for it to be highlighted as a keyword, against that reference corpus. But was the reference corpus reliable concerning *silver*, or was its frequency overrepresented there? I believe the latter to be the case. Doubtful that *silver* has a markedly higher frequency in literary fiction than in other text types, I consulted the Bank of English, which shows that *silver* has a frequency of only 0.0005%, i.e., 50 per million. And if Wmatrix is used to identify keywords in “Bliss”, using its BNC Sampler of Written Imaginative
writing as reference corpus (where silver has a.01% frequency), then silver emerges as one of the story’s top lexical keywords. In the interests of full disclosure I should report that using the BNC Sampler also highlighted another lexical keyword in the story: drawing-room. Evidently there were sufficient drawing-rooms in my own reference corpus for this item to escape attention previously. We shall find drawing-room is of some interest in Section 4 of Chapter 9. Both the silver and drawing-room examples are reminders of the need for vigilance and human review in the inexact science of corpus analysis.

4.2 Clusters

Some disproportionately frequent word sequences, known as clusters or n-grams, are also potentially narratively important – again, if they actually occur. In “Two Gallants”, as we saw earlier, most of the dozen or so 3-word clusters that occur three or more times are quite banal – at least, they feel so for anyone familiar with the story. They include at the corner, the corner of, corner of the, and so on (several of these instances clearly involving alternative segmenting of the same textual sequences). Just one four-word cluster was found to occur as many as three times, namely it off all right, with these immediate co-texts:

I suppose you’ll be able to pull it off all right, eh?
are you sure you can bring it off all right?
he was sure Corley would pull it off all right. All at once

It is noticeable that on each occasion the cluster is accompanied by signals (such as suppose and sure and the question mark) of speaker uncertainty as to whether whatever ‘it’ represents can be achieved. It is also clear from the sense of ‘bring off’ or ‘pull off’ that the task, the ‘it’, entails some difficulty.

In addition to the basic it off all right cluster, it is possible – as shown in Chapter 3 – to link the component words to a little more adjacent text than the triple occurrence of this precise word sequence. This can be done in a fairly mechanical way, e.g. by treating the four-word cluster as comprising (inter alia) two two-word phrases, it off and all right, and seeing if either of these enter into relations with items to their left and right, respectively. A search for other instances of it off in the text reveals a fourth phrase to add to the three cluster instances:

– I’ll pull it off, he said. Leave it to me

Similarly a search for other instances of all right in the text finds an additional four:

– She’s all right, said Corley. I know how
– She was... a bit of all right, he said regretfully.
She'll be there all right. I always let her
– Work it all right now, said Lenehan

Thus, linked to this core of three instances of the *it off all right* cluster we may associate five more partial repetitions or echoes, the network of eight phrases reiterating the idea of pulling something off all right and (and because?) ‘she’ is all right. In terms of textual bulk, the lexical keywords and clusters of the story noted in this section leave a small textual footprint by comparison with keyword-main character sentences or paragraph-initial sentences (Parameters 1 and 3). Despite that small presence, they may be rather important to narrativity and prospection, and may merit special highlighting in the abridgement.
This chapter continues the itemised discussion of narrativity parameters begun in Chapter 7. In that chapter, parameters 1 to 4 were described: sentences containing high frequency keyword character names; sentences with narrative-tense finite verbs in character-depicting action clauses; opening sentences of narrative paragraphs; and sentences carrying lexical keywords and clusters. In this chapter parameters 5 to 8 are presented: sentences carrying characters’ represented thought; prospective directive speech; negated propositions; and projecting modal or mental process verbs. All eight features form a loose grouping, so that the separation into two chapters is chiefly for ease of presentation. Nevertheless, there is a progression from the more ‘atomic’ and overt kinds of signals (single prominent words, particular kinds of verbs, recurrent short phrasal clusters) in the previous chapter to devices with a longer reach, often with a discoursal rather than merely sentential or clausal presence, in the present chapter. Of course even the ‘atomic’ signals, I argue, should be apprehended within their sentential envelope, so that in both chapters the sentence is equally the final unit of analytical attention. But the reason for highlighting sentences and assuming them to be foregrounded for narrativity involves shifting attention now from the single word or phrase to potentially much longer and less delimited textual spans. I have also called the kind of prospection signalling that is discussed in the present chapter ‘embedded’, simply because these four parameters tend to involve subtler textual discriminations, often reflecting the thinking and speaking of the characters (they are mostly ‘characterological’ signals). At the same time, however, I want to insist that these signals are no less overt, in the wording of the text, than the previous chapter’s ‘core’ ones.

The following section explores in detail the possibilities of detecting FIT-carrying sentences in a story using relatively mechanical means. There is widespread recognition of the importance of FIT, to the plot as much as to character-development, in psychologically-oriented modern stories; FIT often occurs at moments of character epiphany or self-discovery. Normally, relative to ambient narration, FIT sequences are narrative high points, especially significant to prospection and progression. I describe a three-stage search procedure which, if it proves reasonably reliable, suggests that FIT is more rapidly detectable via surface
signals than is sometimes supposed. That ease of detection supports a connected assumption: whether consciously or otherwise, readers can hardly avoid noticing ‘during reading’ the FIT segments of a narrative and their deep disclosure of character preoccupations and expectations.

1. The heightened narrativity of characters’ represented thought

Where do all the sentences instantiating the four markers of narrative prospection described in Chapter 7 lead? They are markers that it is tempting to call subliminal but for the fact that they are real textual presences, kinds of relative prominence that a reader cannot simply ignore. Cumulatively, I suggest, they guide the reader to a sense of plotted and expected continuation. But the features cited so far are insufficient – they are still too ‘external’ to character’s motives and crises. What language in the text will most powerfully express these most inward and most critical reactions and developments?

In the modern period the most ‘character-internal’ narrative language, if the story conveys it at all, is the language of represented thought. This may take various forms, all of which stand apart, to a degree, from orthodox narration and directly reported speech: main types include direct thought, free indirect thought (henceforth DT and FIT, respectively) and interior monologue, also known as stream of consciousness. I exclude standard indirect thought (IT) from this grouping on the grounds that it involves an overt refraction or framing by the narrator’s reporting clause, and is generally judged to be used for the relating of less dramatic and more mundane character deliberation. IT is inevitably demoted in this way, by virtue of the very existence of the subtler and more complex FIT alternative. IT notionally reports a character’s thought, but it does not represent it, being too fully cast in the language and orientation of the narrator. DT, FIT, and interior monologue by contrast can all be said to represent thought, and this is why that label is used in this sub-section’s heading.

Here I propose to address only the second of these, partly because DT and interior monologue do not happen to figure in the stories I analyze in this book, and partly because they are typically so extensive in those narratives where they do occur that, notwithstanding their other qualities, they tend to lack the foregrounding and prospective effect of, especially, FIT.

In “Two Gallants” the only forms of palpably character-internal language to be found are neither DT nor extended interior monologue, but free indirect thought.
Most readers and critics identify FIT at just the following five points in the story, with sentences here grouped where in the text they are contiguous:

1. He would be thirty-one in November. Would he never get a good job? Would he never have a home of his own?
2. He might yet be able to settle down in some snug corner and live happily if he could only come across some good simple-minded girl with a little of the ready.
3. Yet it was surely half-an-hour since he had seen the clock of the College of Surgeons. Would Corley do a thing like that?
4. They must have gone home by another way.
5. He knew Corley would fail; he knew it was no go.

These may be regarded as one of the main guides of narrative progression and main shapers of readers’ narrative expectations, if we put some weight on the fact that in these sentences alone Lenehan’s most private reactions and hopes are recorded. They are privileged glimpses of what Lenehan thinks important, and what he expects or hopes for in the immediate future and in the longer term. Clearly, none of this guarantees that events will unfold just as Lenehan might hope; but the reader is entitled to find ‘good cause’ for all these FIT mentions of Lenehan’s hopes and fears, and his alone, if the story does not continue to bear on Lenehan, and on Corley ‘failing’ or not failing. On the basis of Gricean implicature and tellability, the reader assumes there must be a prospection-relevant point to the telling of these private thoughts.

1.1 Automating identification of FIT:

rule 1, narrative modal verbs with pronouns

In any corpus analytic study of FIT, a crucial requirement is to identify some formal means by which, it may be hoped, all and only those passages likely to be character-internal thought may be selected from a text. Such identification has in the past defied reduction to automatic routines (see discussion in Semino and Short 2004: 27–39), and devising a truly general FIT tagger or parser is fraught with difficulties – not least because a mode such as FIT is an ‘in-between’ or ‘anti-’ category, where contextual or pragmatic factors draw the analyst to classify the utterance as neither pure narration nor pure character discourse. There is no set of necessary and sufficient forms for FIT. The category tends to be a semi-stable combination of forms, identified relationally, that is to say by its perceptible contrast with adjacent unambiguous speech or narrative modes. Nevertheless, some progress in isolating the category can be made by proceeding inductively.
In the five FIT passages from “Two Gallants” listed above, co-occurring forms that seem to be diagnostic include the immediately adjacent co-occurrence of a modal verb (would, might, could, and must) and a pronominal Subject denoting a main character or phenomenon (hence in heterodiegetic narrative likeliest to be he, she, it, or they). Furthermore the occurrences in the above passages of modal and pronoun are instances of direct proximity — he would, would he, he might, he could, they must — even though some of the non-inverted examples in principle might have included intervening material such as a comment adjunct: they obviously must have gone, etc. So it may be that an automatic search for, e.g., would immediately preceded or followed by he, she, it, they (then of could, must, might with the same pronouns) may highlight a significant number (but not all) of the sites of ‘high narrativity’ representations of character-internal thought. On the other hand, among the instances listed above, examples (3) Yet it was surely half-an-hour... Would Corley do a thing like that? and (5) He knew Corley would fail both slip through the ‘pronoun + modal’ net. No single search routine can be expected to identify ‘all and only’ the FIT sentences in a narrative.

1.2 Exclude modals following if, whether, that...

Isolating all the narrative (non direct speech) occurrences of he + would and would + he, could + he and he + could, and so on captures a few sentences involving framed thought (IT) along with the more dramatically subjective FIT instances. For example, it picks out this sentence from the text:

(a) He wondered if he had asked her yet or if he would leave it to the last.

Such sentences need to be excluded, being reported thought governed by a framing verb of communication occurring earlier in the same sentence, hence not FIT. At the same time, however, one might well want to retain, among the foregrounded group, both ‘mixed’ forms where a framing verb is followed by wording in the ‘free’ or direct discourse order:

(b) He wondered had he had asked her yet or would he leave it to the last.

and instances where the verb of communication is sentence-final and therefore does not strictly govern the preceding represented thoughts:

(c) Had he asked her yet or would he leave it to the last, he wondered.

The interesting consequence of these quite specific inclusions and exclusions is that, for search purposes, a formal or lexical constraint may be all that needs to be added to the original search rule (rather than a grammatical one making reference to whether the would he is governed by another verb or not). Where the order is he +
would (or, more generally, subject pronoun + modal), the search should go on to exclude cases where, to the left of the phrase and in the same sentence, there occurs any subordinating conjunction such as if, whether, that, etc. This constraint would exclude the actually-occurring sentence (a) above, which is not FIT, while admitting the invented versions (b) and (c), which are partially FIT in form and function.

1.3 Automating identification of FIT:

Rule 2, include all questions and exclamations in the narrative

Additions and refinements to the search seem to enable it to capture four of the five non-contiguous FIT sequences listed above. The sentences under (3) and (5) above are those that escape the first search procedure, based on modals and adjacent Subject pronouns, because in them the modal takes a name, Corley, as Subject rather than a pronoun. Should the selection rule be weakened so as to allow names, as well as Subject pronouns, with modals? Or should the search select all instances of sentence-initial Would occurring in non-direct speech? In “Two Gallants” this would highlight Would Corley do a thing like that? along with the already-identified Would he never get a good job. Would he never have a home of his own? and only those sentences. But a rule of including as FIT all narrative sentences beginning with Would is unlikely to be sufficiently generalisable. To begin with, why limit the rule to sentence-initial Would rather than including Could, Must, Might, Surely, and their negations (e.g., an invented version, Surely Corley wouldn’t do a thing like that?)? Similarly it may be worth noting (also in Sequence 3, above) that the Yet it was surely half-an-hour sentence contains the only occurrence in the story of the disjunct surely, and that when surely occurs outside direct speech in narrative fiction, it is often a signal of FIT.

To address these cases, I propose two further search procedures to the one described above. The second search is for all interrogatives and exclamations in the narrative or non-direct speech part of the text; these should be picked out as putative character-internal thought. It should be possible to make this search, like Search 1, form-based and semi-automatic: simply searching for all sentences in the narrative with a narrative-tense verb and ending with a question mark will identify all interrogatives. For exclamatives the search should cover two patterns: all narrative (non-speech) sentences ending with an exclamation mark; and those with an initial Wh- word, and a narrative tense verb – often BE – that is not immediately followed by a, the, or some. The latter constraints would select How pretty she was as an FIT candidate, with or without final exclamation mark, but perhaps contentiously would exclude sequences such as How pretty was the girl as not well-formed either as an FIT interrogative or as an FIT exclamative.
This second procedure catches sentences like *Would Corley do a thing like that?* Sometimes – and this reflects the ‘undecidability’ of FIT on purely formal grounds, finally – sentences of this type will be attributable to the narrator addressing the reader or a narratee, and not at all judged to be the expression of a character’s thought. For example, *Was that wise, dear reader?* But it is some consolation to note that these are extremely rare in modern stories, and sufficiently marked in their own way to be worth isolating as especially contributory to expectation and narrativity.

1.4 Automating identification of FIT:

Rule 3, include all flanking sentences containing subjective modals

The third and final search procedure is intended to ‘harvest’ sentences like (3) above, *Yet it was surely half-an-hour since he had seen the clock.* As this sentence immediately precedes the *Would Corley…?* one just captured by the second procedure, it might be felt that the former does not need to be entirely independently identified. Exploiting this idea, I propose that a third rule be adopted whereby, for any sentence classified as FIT by Rules 1 or 2, a search is made of the sentences occurring immediately before and after it, for any occurrence of a modal verb or modal adjunct (a ‘core’ list of these could be adopted: *could, would, should, must, might, surely, certainly, probably, possibly, maybe*, and so on). Wherever an adjacent sentence containing one of these items is found, and regardless of presence or absence of negation, that sentence is automatically to be classified as FIT too, i.e., it is presumed to be an extending earlier (or later) of the FIT passage whose nucleus has been identified by Rule 1 or 2.

At this point it should be acknowledged that these three search procedures or rules are incapable of distinguishing free indirect thought (FIT) from its reported speech counterpart, free indirect speech (FIS). Accordingly it might have been more accurate to have referred above, throughout, to the superordinate category that subsumes both FIT and FIS, namely free indirect discourse (FID), and to have presented the search procedures as search routines for the automatic identification of FID. On the other hand, in practice, the stories examined so far tend to use FIT exclusively; and only FIT (and not FIS), with its disclosures of characters’ unspoken thoughts, is a *prima facie* index of prospection.

1.5 Modifying the three FIT-finding procedures

A semi-automatic search for FIT can proceed, it is proposed, by applying the three search procedures listed above to all the non-direct speech material in a narrative text. How well do the procedures fare, when used on a story such as Mansfield’s
“Bliss”? Semi-automatic FIT retrieval by Rule 1, selecting just those narrative sentences in “Bliss” where a personal pronoun and would are immediately adjacent, identifies the following sentences or clauses:

1. Sometimes she [baby] wouldn’t let the spoon go
2. Up to a certain point Miss Fulton was rarely, wonderfully frank, but the certain point was there, and beyond that she would not go.
3. But he [Harry] would pretend to himself that they mattered beyond measure.
4. And then he would make a great point of coming into the drawing-room…
5. It would be silver now, in the light of poor dear Eddie's moon…

Are these all plausibly FIT? In my view (3) to (5) in context are certainly so, all being Bertha’s FIT. (2) is borderline, while (1) certainly is not. Before making any adjustment to the first search procedure, let us consider how it fares in extracting the next batch of putative FIT sentences, those with could and adjacent pronoun:

1. She could not bear the tight clasp of it another moment
2. She couldn’t absurdly cry: “Hasn’t it been a divine day!”
3. Now she was so tired she could not drag herself upstairs to dress

None of these three feels to be fully FIT, so ideally the search criteria should exclude them, although Rule 1 as currently stated would not. But all three in this second group also have two things in common: they have direct (non-inverted) word order, and they are negated (in fact un-negated use of could is very rare in the story). To exclude them, and Sentences 1 and 2 from the preceding list of would examples, we therefore need only to amend the first rule so that it excludes all instances where uninverted subject + modal order is combined with a not or n’t (as in She couldn’t absurdly cry). Note that this amendment is deliberately parsimonious: it leaves inverted clauses (modal + subject: would she…) and other negative constructions (he would say nothing) untouched.

It seems that we also need to exclude examples, like Sentence 3 above, of the form so [adjective/adverb] that she could (not) [verb]. The grammatical criterion of embeddedness or dependency seems important grounds for excluding Sentence 3, although it would be excluded now anyway by the ‘pro + modal + not/n’t’ amendment just introduced. It is standardly recognised in grammars that could may be used to express ability or possibility; many of the examples just excluded concern ability, whereas the could of FIT is usually expressing the focalised character’s mental exploring of possibility.

A further remark is in order concerning Sentences 1 and 2 of the would examples above, repeated here for convenience:

1. Sometimes she [Baby] wouldn’t let the spoon go
2. Up to a certain point Miss Fulton was rarely, wonderfully frank, but the certain point was there, and beyond that she would not go.
As noted, the 'pro + modal + not/n't' condition summarily excludes these sentences from classification as FIT by Rule 1. While at first this looks too crude and simple, it leads to fairly satisfactory results when all three searches are used in sequence. To begin with, we certainly do want Sentence 1 excluded from the FIT set. Most interesting is the excluding of Sentence 2 by Rule 1, which turns out to be only temporary, since the sentence that immediately follows it in this story is the following:

Was there anything beyond it?

Now that sentence will be independently selected by Rule 2, and this allows the 'borderline' Sentence 2 of the would sentences, about Miss Fulton as wonderfully frank, to be selected by Rule 3, since it contains a modal. But being a Rule 3 selection, the sentence is as it were explicitly marked as borderline. This and subsequent examples require that rule-ordering is observed: Rule 2 must apply after Rule 1 and may override it, not the reverse; and Rule 3 applies only after both 1 and 2 have been applied. Incidentally, if Rule 3 were amended to be allowed to apply recursively, then the extent of identified FIT passages could enlarge, potentially; for example, Sentence 3 of the could examples excluded above could be reinstated as FIT potential (it occurs in the text adjacent to a Rule 3-identified FIT sentence). But as presently stated the rule does not allow this.

Rule 1 has now been modified from being a simple selection wherever there is (in the narrative text) a modal with immediately adjacent personal subject pronoun, so that it now excludes direct-order instances with following not/n’t, and all instances preceded by subordinators such as if, whether, because, although, and so on. Application of it with respect to other modals in the narration of “Bliss” proceeds quite smoothly. Thus there is only one instance of must + pro in the narration, and the clause is easily classifiable as FIT:

It must have been the spring.

There is only one instance of should + pro, also, and while this is potentially FIT it is excluded here as occurring in an embedded clause. In the more mechanical terms amenable to corpus searches, it is excluded as it is preceded in the same sentence by the subordinator how:

What she simply couldn’t make out—what was miraculous—was how she should have guessed Miss Fulton’s mood so exactly and so instantly.

However this sentence will be classified as FIT on a second cycle by the same means that applied above to the Miss Fulton was wonderfully frank sentence. Again this is because the next sentence in the text (For she never doubted for a moment that she was right, and yet what had she to go on?) is classed as FIT by Rule 2 (it is a narrative sentence interrogative), and this enables the adjacent What she simply
couldn't make out...sentence to be classed as FIT by Rule 3. These examples underline the importance of ordered application of the three selection rules.

Are there, on the other hand, any sentences in the narration using could, would, etc. which may be FIT but that are not caught by Rule 1 (amended to exclude subordinator-embedded and direct-order not/n’t-carrying instances), since they involve no modal-adjacent personal pronoun? I could identify only one:

What was there in the touch of that cool arm that could fan–fan–start blazing–blazing–the fire of bliss that Bertha did not know what to do with?

But this is immediately satisfactorily caught by Rule 2. Indeed there are a number of narration-section interrogative and exclamative sentences in the story, captured as putative FIT by Rule 2, as specified above. They include the following:

Was this what that feeling of bliss had been leading up to?
What a boy he was in some ways–so impulsive–so–simple.

Nor do there seem to be any sentences ill-advisedly caught in the FIT net by Rule 2. Several sentences in the course of the story, including the second and third ones at its opening (What can you do...? Oh, is there no way you can express it?), which are plausibly classed as DT, will be excluded because of their non-narrative tense. Incidentally, to capture verbless FIT sentences like And the others? it may be best not to require that the FIT interrogative carry the narrative tense (here, past), but only that it does not carry an alternative tense.

In summary, I have proposed a fairly crude and simple set of form-based diagnostics on the basis of which particular sentences in a short story can be set apart and considered as possibly representing a character’s inner thoughts. These diagnostics do not capture all instances of FIT; nor can they entirely reliably indicate the scope, extending to adjacent sentences, of a FIT passage; nor can they identify all the traces and incipient FIT colourings sometimes apparent in predominantly narratorial sentences. But they are sufficiently reliable to be worth persisting with. The three-stage search fails to capture just one attested FIT sentence in “Two Gallants”: He knew Corley would fail; he knew it was no go. That and similar sentences are the focus of a separate prospection parameter, discussed in Section 8 below.

Both the main procedures and the third, contingent one may be relatively easy to adapt to the concordancing and collocating programmes of WordSmith Tools, Wmatrix, or similar – and can even be explored (more laboriously) with the simple Search function in most word processing packages. The first procedure is straightforward unless it emerges that an author has used a less core modal, such as HAVE TO: They had to have gone home by another way. The second search involves, for all interrogative sentences, scrutiny of all non-speech sentences ending with a question mark which also have inversion (in FIT this will usually – perhaps overwhelmingly
often – take the form of a personal subject pronoun preceded by a modal or one of
the set do/does/did/have/has/had/am/is/are/was/were, with or without an interven-
ing negation marker). For exclamative sentences in the second test the search will
be for simple sentences beginning with a WH-word but without inversion of (usu-
ally) the same pronoun and modal or operator verbs just listed.

The three-rule FIT automatic searcher can be applied to a variety of stories
and longer fictional narratives, but unsurprisingly the less a story resembles the
Joyce and Mansfield ones that were the ‘testbed’ for the model, the less reliable the
model is likely to be. A determination of the limits of the model’s usefulness and
reliability awaits further testing on a diversity of literary narratives. Initial find-
ings give some cause for optimism. The three-stage procedure works reasonably well
on Woolf’s “Mrs Dalloway in Bond Street”, on Kipling’s “Mary Postgate”, on Hem-
ingway’s “Cat in the Rain”, and on Joyce’s “Eveline”. On a story such as Carver’s “A
Small, Good Thing” (in general Carver makes scant use of FIT – see Bramlett and
Raabe 2004), results are more mixed: the procedure automatically selects most of
the FIT and FIS, along with some sentences which convey a particular character’s
focalisation but not their wordings, being, rather, what Fehr (1938) long ago called
substitutionary perception.

Further applications will undoubtedly reveal ways in which the rule-state-
ments may need adjustment. Already two secondary potential adjustments can be
noted as likely to help exclude non-FIT sentences. Both adjustments concern the
first procedure: it might be desirable to restrict candidate pronouns to singular
ones; and it might be appropriate to discount any sentences where pro + modal is
followed at once or proximately by have.

As the three-rule FIT-tester is applied to a larger range of texts, it becomes
newly clear – as all the authoritative surveys of the topic tell us – that there is no
generalisable means of distinguishing FIT from FIS on formal grounds. Therefore,
as mooted earlier, the repertoire of probe procedures as described might be more
accurately described as a FID-finder. Generally, it will be only on the grounds of
co-textual indications, such as presence in the context of another character and
textual inclusion of their reported interlocutions, that a determination between
FIS and FIT status can be made. In other words if a FID-identified sentence like
Would Corley do a thing like that? were preceded or followed by speech reporting
to which it appeared to be oriented, then FIS can be assumed; otherwise, FIT may
be assumed. Perhaps in the 20th century FIT is the default mode, of these two.
Would he never get a job? may be interpreted as FIT except where, for example, it
is preceded by The girl asked him if he would ever make anything of himself; or is
followed by “Of course you will!”, his friend assured him. Thus simple formal clues
to a FID sentence’s FIS status will include proximate speech marks and inquit
verbs. Here, ideas emerging from cognitive narratology (e.g., Jahn 1997; and
Palmer 2004) may be of help. Palmer, for instance, in studies of how readers’ use their world-knowledge in interpreting fictional characters’ represented minds, attends to a variety of textual signalling.

2. Prospective direct speech

Using the five factors discussed so far in order to ‘reduce’ the text to a concentrated abridgement carrying the narrative’s prospection-significant material, the most glaring neglect has been of direct speech. Characters’ reported direct speech is a vital means of advancing many stories; a fuller attending to it requires no lengthy justification here. And yet, so far, this mode of speech presentation has been included in the identification of highly prospective text only on some more general ground or a contingent basis. Thus in “Two Gallants”, for example, direct speech has only been classed as high-narrativity material where it happens to come in a sentence in which the keyword Corley or Lenehan occurs in the reporting clause. At the same time it would be implausible to suggest that all direct speech was especially relevant to prospection and progression. Assuming that we can make reliable formal discriminations, and that only certain kinds of direct speech are particularly guiding of reader expectation, we need to distinguish those kinds carefully and, again, preferably in ways that allow semi-automatic corpus identification, using a simple and reasonably reliable form of instrumentation.

What do people do with speech? Following Halliday, I have argued elsewhere for a simplest schema of speech moves comprising four main initiating transactional types – informs, questions, requests, and undertakings (TooIan 1998, 2000a, 2000b). To these we need to add at least a fifth category, and arguably several more, to identify and distinguish the various interactional and reacting moves performed through turns of speech or text-messaging: phatic remarks, acknowledgements, apologies, compliments, jokes, insults, ‘chat’, and so on. It would be an overstatement to say that none of the latter interactional types of move prospect at all; but I will proceed under the assumption that the four transactional moves are more prominent in creating event prospections in whatever discourse they figure in. The prospective potential of direct speech informs, questions, requests and undertakings is sketched in the following paragraphs.

Within direct speech, perhaps all utterances functioning as questions should be treated as important for prospection. In all but rhetorical situations, they create a local prospection (of an answer or a deflection); they also standardly articulate a segment of the story world (of events, facts, etc.) of which the speaking character has no more than partial knowledge, with the expectation of disclosure of fuller knowledge in the interlocutor’s response. Because a character’s question articulates their interests and limited knowledge, special reader-attention to each question is
arguably a necessary part of empathetic alignment, of feeling-with a character, i.e. experiential reading.

By the same token direct speech which carries any sort of directive or undertaking thereby prospects particular future actions or developments in the story, rather than being oriented to the past. Such directives and undertakings can give rise to commensurate reader expectations, large or small. So these too merit particular attention. What remains? All the direct speech which is notionally ‘non-transactional’ but only interactional, such as compliments, apologies, phatic remarks, casual conversation of the kind that evades a staged structural description (Eggins and Slade 1977); and all the instances which are purely acts of informing.

The most direct textual signalling of questions and directives in direct speech are such features as the question mark, with or without pronoun-verb inversion. For undertakings, a standard textual clue is I followed by will or other modal, or other means of conveying future relevance (simple present, present progressive, with future-time-reference phrase: e.g. I am going to repay you tomorrow). None of these markers is an entirely automatic indicator, however. For instance it is implausible to classify Corley’s Didn’t she tell me herself? as a genuine question; it is merely a rhetorical one functioning as an emphatic (past-oriented) inform.

The inform category remains very extensive and where the inform is oriented to the future in the story world, it can be potentially just as relevant to narrative expectations as an undertaking or a directive. It differs from them in that it is a prediction about a coming change for which the speaker is not directly responsible (as they are in any kind of undertaking). Thus future-oriented informs are in several respects close semantically to undertakings, but without speaker as Agent. Again, then, direct speech using the future-oriented modals (and most of them can be used with future-orientation: not just this war will end but also this war should/ must/ might/ may/ ought to end) can be highlighted along with all instances of questions, directives and undertakings.

Questions, directives, undertakings, and future-oriented informs are important to prospection because they entail a sequel – even if no sequel occurs or if what follows seems incomplete as a response. They are classifiable as either local or long-distance initiations, even when they also function as an encapsulation or completion of prior text. Because they are sequel-prospecting initiations – “first pair parts” in conversational analytic terms – there seems a clear justification for noting, in this reducing and foregrounding exercise, any material that can be reasonably interpreted as their sequel or second pair part. An example of one such
simple and local prospection and completive response pair is Lenehan’s exchange with the serving-girl in the shop:

- How much is a plate of peas? he asked.
- Three halfpence, sir, said the girl.

The reader is entitled to assume that Lenehan’s question has been fully answered, that it excites no longer-term expectations concerning the direction of the narrative. This might also enable the analysis to distinguish between prospection-generating moves which seem satisfactorily answered and whose prospective force is thereby in a sense cancelled, and those questions, directives, and so on that remain open or in force, noticeable for not (yet) being resolved in the text. Each prospection-initiating speech move could be conceived of as a lefthand or opening bracket, logically linked to a later closing bracket whose occurrence is projected, even if it fails to be realised in the text.

In the first instance, however, I propose highlighting and focussing on just the prospection-openers and not on any completive sequels, as points of high reader expectation, and notwithstanding the asymmetry this creates. My reasoning is partly that our interest is in all the resources that trigger prospection, and not those that seemingly cancel or discharge the prospection. But it is noticeable also that well-formed answers often do not have the power fully to meet and erase the prospection in an initiating question. Thus when Lenehan asks Corley about his woman friend:

- Well... tell me, Corley, I suppose you’ll be able to pull it off all right, eh?

this is ‘answered’ by:

Corley closed one eye expressively as an answer.

Subsequently Lenehan pursues this:

- Is she game for that? asked Lenehan dubiously.

and gets this reply:

- She’s all right, said Corley. I know the way to get around her, man.

But something remains prospected by Lenehan’s questions. They can only be fully answered when we find out whether or not the woman is ‘game’ in actuality, and not just in prediction. While both Lenehan’s questions are at least implicitly answered in the complacent affirmative by Corley, the questions themselves prospect some sort of future transaction (it and that in the respective questions). The transaction’s precise nature remains obscure here, for the reader in mid-story; but that reader can now expect to witness or learn more about Corley (trying to) pull it off.
So while Lenehan’s questions to Corley about a future undertaking are answered, that prospected undertaking remains to be fulfilled.

3. Negation-carrying clauses

Treating all negation-carrying clauses in the narration as a kind of narrative foregrounding is to endorse Gerald Prince’s influential category of the ‘disnarrated’: “terms, phrases, and passages that consider what did not or does not take place” (Prince 1988: 3). Disnarrated material in a story is as fascinating and strictly unnecessary as repeated material, and often contributes to prospection by disclosing what almost happened or might have been the case. Prince’s colleague at Pennsylvania, the linguist William Labov (1972), has equally influentially included negation among the most basic evaluative resources of oral narratives, again because any negative narrative clause – *He didn’t jump out of the car, instead he…* – projects a possible narrative continuation for us fleetingly to imagine and puts it in evaluative contrast with the averred actual continuation: *instead he put the accelerator to the floor and just flew across the gap.*

Part of the reason for looking at negatively cast narration is that, as Nørgaard puts it, “the negatives in fact create what they negate” (Nørgaard 2006: 41; see also Werth 1999 and Hidalgo Downing 2002). They evoke a possibility or condition and simultaneously deny it, so are a vivid (and deeply grammaticised) instance of Bakhtinian polyphony, meaning that “more voices are at play in the text” (Nørgaard 2006: 48). Nørgaard discusses the example from “Two Gallants” where we are told that “He [Lenehan] sat down at an uncovered table.” The negation *uncovered* not only inevitably alerts us to a condition (the table being covered) which does not here apply; it also implies that in the general circumstances a covered table might have been expected or would have been more normal or standard. The uncovered table is implicitly different from the norm or standard (it is barer, poorer, more humble). So negations not only project two perspectives, but also often imply that the negated perspective is the more ordinary, reasonable and expectable and that to correct what the reader might otherwise expect, its absence is being reported. The only other direct mention of coverings in “Two Gallants” comes in the harp-playing episode, where the harpist’s instrument is invested with strange female personification: *His harp, too, heedless that her coverings had fallen about her knees.* There is a kind of double covert negation here: that the coverings had fallen about the harp’s “knees” is evidently an uncovering, and it is one that “she” is heedless of. Again the implication is that none of this is entirely normal, ordinary, or appropriate.
How wide should analysts cast their net when trawling a text for markers of negation? And how much of the co-text should be treated as foregrounded along with the negative particle, by virtue of the latter’s presence? Is the scope of contextual influence different when a negated adjective is present, for example, than when a narrative-tense verb is negated? In my story abridgments, I have adopted the practice of selecting each marker of negation along with the full clause to which it relates. The markers of negation searched for include the following: not, n’t, never, no, nothing, nowhere, fail (to), -less, non-locative without, dis-, un-, im-, and in-. This policy means that in a sentence like John told Mary that he was unhappy that she was leaving the company, I would select he was unhappy that she was leaving the company, and not merely he was unhappy, as the scope of the negation-marker un-. Although one might expect the search to be laborious and protracted, this seems to be rarely the case: in “Two Gallants” as a whole for example (i.e., including the direct speech that is outside the interest of this parameter), there are only ten instances of not, and five of n’t.

To simplify and reduce the body of data for consideration, it is tempting to reflect the distinction between lexicalised and merely morphological negation, and to assume that the lexically autonomous negations (especially not and its contraction n’t, no, never, nothing, nowhere, fail (to), and without) are more foregrounded and narratively significant than those words that simply incorporate a negative derivational morpheme (-less, dis-, un-, im-, and so on). Thus he laughed noiselessly for fully half a minute would by that criterion be grouped among the less prominent reversals of a perspective projected but denied (in which he laughed noisily). These arguably do not attract special reader attention in the way that ones with lexicalised negation, such as No one knew how he achieved the stern task of living, do. An advantage of privileging free negative morphemes would be that the search could be relatively automatic, whereas searching for words with in- and un-negative prefixes requires some assistance (such as Parts of Speech tagging) to ensure the exclusion of words such as into, under, uncle and ink. On the other hand enforcing this distinction would cause us to ignore he laughed noiselessly but include its (non-occurring) near-paraphrase he laughed without making a sound. On balance, even though the body of data is thereby enlarged, it seemed safer as an initial policy to treat as foregrounded every clause in which there was either a lexical or a morphological negation. Interestingly, in “Two Gallants” the resulting expansion of the data is slight: there are surprisingly few negative suffixes in the narrative text – for example only one instance of the negative prefix in-, insensitive. Finally we must accept that there will always be borderline words, arguably negative but with little in them that perceptibly renders them negative in grammar or morphology, which the search will miss. “The Love of a Good Woman” alludes at its opening to the anonymous donor: paraphrases of anonymous highlight its
negative element, but does the word carry a marker of negation? Historically, the form combines the ancient Greek negative prefix *an* with *onoma* (name), but since *an* is not a psychologically real negation-carrier for most modern readers, *anonymous* (and *anaerobic, anabolic, anisomorphic*, etc.) will go unharvested. But the example may also amount to evidence in support of the idea that there are degrees of negation in English grammar, and that ultimately the most interesting specifically with regard to narrative progression are those conveyed by free morphemes (such as *not, never, fail to, without, nothing, and nobody*).

4. **Narrative verbs of modality and mental processing**

So far, modality has not been directly addressed in these chapters’ array of foregrounded carriers of narrative progression. Modality clearly plays a prominent role in such parameters as FIT, future-oriented direct speech, and perhaps even in negated sentences. But modality viewed independently has not been proposed as a main force in the creation of expectations as to how a story, and especially its plot, will continue. While it is powerfully evaluative, crucial to the disclosure of subjectivity and the delineation of character, it is not in itself, divorced from other effects such as FIT, consistently central to prospection.

And yet when they operate alongside other linguistic resources, modalised qualifications in the narration can be highly revealing. In his corpus stylistic study of *Heart of Darkness*, Stubbs suggests that verbs are of stylistic interest, in particular those that ‘concern uncertainty, perception and knowledge’ (2005: 11f); and elsewhere he has commented that *say* and mental process verbs are frequent in fiction in general (Stubbs and Barth 2003). We can see as much in a simple narrative, such as H. C. Andersen’s story of “The Princess and the Pea”. The prince wants but is uncertain about how to find a real princess, the princess arrives, says she is a real princess, the old queen plans but says nothing, she puts a pea in the bed, the princess sleeps dreadfully, so that now they knew she was a real princess: modalised verbs (and a few states, such as being royally hypersensitive) drive the narrative. We also saw in Section 1 of this chapter that the single FIT sentence in “Two Gallants” that the three-stage search-routine failed to identify was *He knew Corley would fail; he knew it was no go*. If the FIT-finding procedure misses this important suspense-furthering sentence, another of the prospection-highlighting methods perhaps should contrive to catch it.

A detailed study of all the narrative past tense verbs in “Two Gallants” reveals that on several occasions they report Lenehan’s mental processes, in a form that is often what Leech and Short (1981 [2007]) call a narrative report of a thought act (NRTA) rather than FIT. Despite being ‘only’ NRTA, these intuitively feel important
enough to plot and characterisation that we might wish to have them included in a prospection abridgement (even if, for consistency, the harvest would have to include those with Corley or others as Senser). The relevant verbs include thought, knew, and wondered:

*He found trivial... He knew that he would have to... He thought how pleasant... He knew what those friends... he knew the girls... He wondered if he had asked her yet.*

Where these verbs introduce an entire proposition, they are often particularly informative about the thinking or reacting character. Accordingly, within the listing of a story’s narrative verbs (extracted by using Wmatrix’s VVD category, or similar tagging software), those mental process or modalising verbs that represent kinds of knowing or desiring are treated as meriting special attention, as central to an eighth and final narrativity parameter. The relevant modalising verbs include know, think, seem, appear, suspect, expect, want, need, see, look, wonder, believe, and realise, among others. The verb seemed, for example, occurs as many as 10 times in “Bliss”, and this prompts the question whether special attention should be paid to its particular contexts of use. Most of the verbs just listed are called, in systemic linguistic description, verbs of mental processing. Even those that would be classified differently, such as seem and appear, strongly imply a mental perspective from which the judgement they carry is made: the narrator’s as a default, but a particular character’s where the co-text justifies assuming such a restriction in the implied source of assessment.

A crucial feature of mental process verbs is that they can project an entire following proposition. It is only when a mental or modal verb does project in this way that the sentence in which it appears will be classed as an instance of the eighth parameter. Thus where in “Two Gallants” we are told of Corley that “He knew the inner side of all affairs”, this is passed over, being judged to be of lesser relevance to narrativity than being later told, of Lenehan, that “He knew Corley would fail; he knew it was no go”, where on both occasions knew projects a full clausal proposition. Similarly modal verbs such as seem and appear will only be of interest where they participate in the projection of an entire clausal proposition, rather than functioning intensively in a relational process (such as And it had seemed quite sense at the time, in “Bliss”).

In this area of textualised subjectivity it is especially clear that a simple corpus-oriented searching cannot capture all the metaphorised or relexicalised textual expressions of mental or modal subjectivity, where verbal resources other than the standard mental and modal verbs are deployed. Consider for example the report from late in the “Two Gallants” story, that An intimation of the result pricked him [Lenehan] like the point of a sharp instrument. This is vividly expressive of Lenehan’s expectations but uses no standard modal or mental verb, so would not be
caught and included under this final parameter. The sentence uses a formulation involving grammatical metaphor (Halliday and Matthiessen 2004) that is far removed from the congruent and codified (such as *He suddenly sensed that the venture had failed*). As it happens, in this textual instance the ‘projected’ thought (*An intimation of the result*) is also phrasal rather than propositional, so would be excluded on that ground also.

It is relatively easy to isolate all those narrative sentences containing one of a short inventory of mental or modal verbs in the narrative tense, where either a named main character or a personal pronoun is adjacent (e.g., in “Bliss,” *Bertha, she, Harry, he, it, they*). Some of these sentences, those containing a frequent-keyword character name, will have already been selected for the abridgement. Others would be by this means highlighted for the first time. But it should be acknowledged that in venturing into this area of the modalising of narration one is turning to the more subtle, complex and delicate aspects of narrative texturing, and that this is not an unqualified advance. Whatever patterning there is in this area is perhaps necessarily less regular and overt, and certainly less amenable to rapid electronic retrieval via corpus methods. Its subtlety of effect also suggests that it is less immediately directing of the reader, and less psychologically plausible as an *inescapable* influence on readers as they form narrative expectations in the real time of reading. In contemplating narrative sentences containing projecting mental and modal verbs the analyst is moving to the textual borderland where signalling of progression is intermingled with signalling of evaluation.

For ease of reference, a tabular summary of the eight proposed sentential types judged central to narrative prospection is provided as Figure 4 at the end of this chapter.

5. Implementing the model with “Two Gallants”

The remaining sections of the chapter present and comment further on the selective version (or abridgement) of the “Two Gallants” story that emerges when sentences carrying one or more of the eight types of prospection cue discussed above are extracted from the text and treated as the story’s foregrounded seam of prospective material. Selecting just those sentences (for some parameters, clauses) of the text in which one or another of the postulated narrativity cues occurred, and de-selecting all others, the full text of 3,900 words is reduced to one of 1,692 words, i.e. approximately 40% of the original. This is reproduced below, following a Key to the graphological means used to distinguish the eight types of prospection foregrounding, and then a few comments follow on how the abridgement was prepared, with discussion of the main difficulties.
Key

Bold: parameter 1, high-frequency keyword character names (here, Corley and Lenehan), within (unbolded) their sentential co-texts.

Underlining: parameter 3, first sentence of paragraph.

Italics: parameter 4, lexical keywords (walk* and street*) and key clusters (it off all right).

Bold wavy underlining: parameter 5, automatic/mechanical identification of potentially FIT sentences.

Broken underline: parameter 6, future-oriented (prospective) direct speech (requests, undertakings, etc.)

Shadowed: parameter 7, negation-carrying words, phrases or clauses

Single wavy underlining: parameter 8, projecting modal and mental process verbs

NB parameter 2, narrative tense verbs, is not independently searched for or listed (see discussion in s.2.3 of Chapter 7). Only those instances of parameter 4 are noted here where they occur in a sentence selected on the basis of another parameter.

The grey warm evening of August had descended upon the city and a mild warm air, a memory of summer, circulated in the streets. The living texture below which, changing shape and hue unceasingly, sent up into the warm grey evening air an unchanging unceasing murmur.

Two young men came down the hill of Rutland Square. He laughed noiselessly for fully half a minute

He became serious and silent when he had said this. Most people considered Lenehan a leech but, in spite of this reputation, his adroitness and eloquence had always prevented his friends from forming any general policy against him. He was insensitive to all kinds of discourtesy. No one knew how he achieved the stern task of living.

– And where did you pick her up, Corley? he asked.

Corley ran his tongue swiftly along his upper lip.

– Maybe she thinks you’ll marry her, said Lenehan.

– I told her I was out of a job, said Corley.

Lenehan laughed again, noiselessly.

Corley’s stride acknowledged the compliment. Corley was the son of an inspector of police and he had inherited his father’s frame and gut. He spoke without listening to the speech of his companions.

Lenehan offered his friend a cigarette. As the two young men walked on through the crowd Corley occasionally turned to smile at some of the passing girls but Lenehan’s gaze was fixed on the large faint moon circled with a double halo.
Well... tell me, Corley, I suppose you'll be able to pull it off all right, eh?

Corley closed one eye expressively as an answer.

Is she game for that? asked Lenehan dubiously. You can never know women.

She's all right, said Corley.

You're what I call a gay Lothario, said Lenehan.

But Corley had not a subtle mind

There's nothing to touch a good slavery.

By one who has tried them all, said Lenehan.

First I used to go with girls, you know, said Corley, unbosoming: girls off the South Circular.

as if he was conscious of being disbelieved

But Lenehan could well believe it; he nodded gravely.

And damn the thing I ever got out of it, said Corley.

Ditto here, said Lenehan.

Only off of one of them, said Corley.

She was... a bit of all right, he said regretfully

I suppose that's your doing, said Lenehan.

There was others at her before me, said Corley philosophically.

This time Lenehan was inclined to disbelieve

You know you can't kid me, Corley, he said.

Honest to God! said Corley. Didn't she tell me herself?

Lenehan made a tragic gesture.

As they passed along the railings of Trinity College, Lenehan skipped out into the road and peered up at the clock.

Time enough, said Corley. She'll be there all right. I always let her wait a bit.

Lenehan laughed quietly.

Ecod! Corley, you know how to take them, he said.

I'm up to all their little tricks, Corley confessed.

But tell me, said Lenehan again, are you sure you can bring it off all right?

His bright, small eyes searched his companion's face for reassurance.

Corley swung his head to and fro as if to toss aside an insistent insect, and his brows gathered.

I'll pull it off, he said. Leave it to me, can't you? Lenehan said no more. He did not wish to ruffle his friend's temper, ... and [be] told that his advice was not wanted.

But Corley's brow was soon smooth again.
They walked along Nassau Street and then turned into Kildare Street. Not far from the porch of the club a harpist stood in the roadway, playing to a little ring of listeners. He plucked at the wires heedlessly.

The two young men walked up the street without speaking. The mournful music following them.

– There she is! said Corley.

At the corner of Hume Street a young woman was standing. Lenehan grew lively.

– Let's have a squint at her, Corley, he said.

Corley glanced sideways at his friend and an unpleasant grin appeared on his face.

– Are you trying to get inside me? he asked.

– Damn it! said Lenehan boldly. I don't want an introduction. I'm not going to eat her.

– O... A look at her? said Corley, more amiably. Well... I'll tell you what. I'll go over and talk to her and you can pass by.

– Right! said Lenehan.

Corley had already thrown one leg over the chains when Lenehan called out;

--And after? Where will we meet?

– Half ten, answered Corley, bringing over his other leg.

--Where?

--We'll be coming back.

– Work it all right now, said Lenehan in farewell.

Corley did not answer. Without saluting, [he] began at once to converse with her:

Lenehan observed them for a few minutes. The ends of her tulle col-

larette had been carefully disordered. Frank rude health glowed in her face, on

her fat red cheeks and in her unabashed blue eyes. As he passed Lenehan took off his cap and, after about ten seconds, Corley returned a salute to the air.

Lenehan walked as far as the Shelbourne Hotel where he halted and waited. As he walked on slowly, timing his pace to theirs, he watched Cor-

ley's head which turned at every moment towards the young woman's face like a big ball revolving on a pivot.

Now that he was alone his face looked older.

He walked listlessly round Stephen's Green and then down Grafton Street. He found trivial all that was meant to charm him and did not an-

swer the glances which invited him to be bold. He knew that he would
have to speak a great deal, to invent and to amuse. The problem of how he
could pass the hours till he met Corley again troubled him a little. He
could think of no way of passing them but to keep on walking.

He was hungry for, except some biscuits which he had asked two
grudging curates to bring him, he had eaten nothing since breakfast-time.

He sat down at an uncovered wooden table...

- How much is a plate of peas? he asked.
- Bring me a plate of peas, he said, and a bottle of ginger beer.

He spoke roughly in order to belie his air of gentility for his entry had
been followed by a pause of talk. When he had eaten all the peas he sipped
his ginger beer and sat for some time thinking of Corley's adventure. In
his imagination he beheld the pair of lovers walking along some dark road;
he heard Corley’s voice in deep energetic gallantries and saw again the leer
of the young woman's mouth. He would be thirty-one in November. Would
he never get a good job? Would he never have a home of his own? He
thought how pleasant it would be to have a warm fire to sit by and a good
dinner to sit down to. He knew what those friends were worth. But all
hope had not left him.

He paid twopence halfpenny to the slatternly girl and went out of the
shop to begin his wandering again. His friends asked him had he seen
Corley and what was the latest. He replied that he had spent the day with
Corley. At this Lenehan said that he had been with Mac the night before
in Egan’s. Lenehan did not know:

He left his friends at a quarter to ten and went up George's Street. He
set off briskly along the northern side of the Green hurrying for fear Cor-
ley should return too soon. He leaned against the lamp-post and kept his
gaze fixed on the part from which he expected to see Corley and the young
woman return.

His mind became active again. He wondered had Corley managed it
successfully. He wondered if he had asked her yet or if he would leave it to
the last. But the memory of Corley's slowly revolving head calmed him
somewhat; he was sure Corley would pull it off all right. All at once the
idea struck him that perhaps Corley had seen her home by another way
and given him the slip. Yet it was surely half-an-hour since he had seen the
clock of the College of Surgeons. Would Corley do a thing like that? They
must have gone home by another way.
Suddenly he saw them coming towards him. They were walking quickly, the young woman taking quick short steps, while Corley kept beside her with his long stride. He knew Corley would fail; he knew it was no go. They turned down Baggot Street and he followed them at once, taking the other footpath. Corley remained standing at the edge of the path, a little distance from the front steps. Corley turned and went towards her. The door closed on her and Corley began to walk swiftly towards Stephen’s Green.

Lenehan hurried on in the same direction. He called out: Hallo, Corley! Corley turned his head to see who had called him, and then continued walking as before. Lenehan ran after him, settling the waterproof on his shoulders with one hand. Hallo, Corley! he cried again. He came level with his friend and looked keenly in his face. He could see nothing there. Well? he said. Did it come off?

They had reached the corner of Ely Place. Still without answering, Corley swerved to the left and went up the side street. Lenehan kept up with his friend, breathing uneasily. Can’t you tell us? he said. Did you try her? Corley halted at the first lamp and stared grimly before him.

6. Compiling the abridgement

Compiling of the abridgement began with extraction of the Corley and Lenehan keyword sentences. These alone comprise the bulk of the abridgment (1,100 words), but they are sentences which often also contain one or more of the other narrativity parameters. As proposed earlier, sentences reporting main protagonists’ actions using narrative tense verbs (prospection Parameter 2) were not independently searched for. This was because in practice to a large degree they are captured in the course of more specific parameter searches. That is to say, the character-action verbs tend to occur in sentences which also contain a frequent character-naming keyword, or are negation-carrying, or paragraph-initial, etc. A brief further comment on this strategy is made towards the end of this section.
The original story was then examined using the FIT/character-internal thought identification criteria (Section 1.5), and all the well-attested FIT sentences in the story were captured, save one: He knew Corley would fail; he knew it was no go. That sentence is already in the abridgement, however, since it includes the top keyword Corley; and, as we more recently saw, it will be captured again by Parameter 8. Equally importantly, the search harvested only two non-FIT sentences, which are therefore included in the abridgement, like the other semi-automatically-identified FIT sentences. They would independently qualify for the abridgement in any case, as they contain negative morphemes. But even in terms of the Parameter 5 focus on the special relevance of characters’ internal thought they are of interest. This is because they are instances of NRTA and substitutionary perception respectively:

He could think of no way of passing them but to keep on walking.
He could see nothing there.

They are not so fully inward of the character as to disclose the exact wording, save for tense and pronoun adjustments, of his thoughts (as FIT purports to do); but they report in summary his thinking and perceiving.

It is interesting to find that some of the FIT-identified sentences were already selected as top-keyword sentences, thus in effect that they are on two independent bases foregrounded for expectation and narrativity. But it also noticeable and unsurprising that some FIT sentences were not sentences containing either of the keyword proper names: on the contrary, a pronoun rather than a name, and processes of self-assessment where mention of another proper name is less likely, are the norm in FIT. The upshot is that while the decision to focus first only on those sentences where Corley and Lenehan are named and not pronominalised tends to undervalue the cohesive chains of pronominal naming and the sentences in which these occur, attention to pronoun-using FIT sentences partially corrects that imbalance.

Section and paragraph opening sentences were now added (Parameter 3), where not already harvested: these sentences are underlined in the abridgement. As noted, it is sometimes a matter of judgement as to whether a sentence (e.g., one of two in a paragraph embedded in a long passage of direct speech dialogue) is genuinely paragraph-initial, particularly where there is a formatting convention that at termination of one turn of talk, whatever follows (speech or narrative) should be preceded by an indentation. Here, just the first sentences of the narrative paragraphs of the story were included.

Sentences featuring the 7th parameter, being negation-carrying, were also often independently selected on a separate criterion. The repeated finding that many of the abridgement sentences display two or more of the eight nominated narrativity features seems to justify selecting for these specific features in the search for key
expectation cues. It also seems confirmatory of the notion that crucial narrativity involves a concentration – even a redundancy or ‘overcoding’ – of cues or triggers, analogous to a symphonic composition’s use of all sections of the orchestra at certain musical high points.

A minor consequence of the addition of several of the negation-carrying sentences that were not on other grounds already selected is worth noting. This is that on many occasions they tend to ‘smooth out’ semantically certain very awkward (incoherent) transitions in the top-keywords abridgement, top keyword sentences remaining the source of most of the extracted text. Thus prior to addition of negation-sentences to the abridgement there is this sequence:

Corley swung his head to and fro as if to toss aside an insistent insect, and his brows gathered.
Lenehan said no more. But Corley’s brow was soon smooth again.

After the addition of intermediate negation-sentences (here underlined), the sequence is much smoother:

Corley swung his head to and fro as if to toss aside an insistent insect, and his brows gathered.
Lenehan said no more. He did not wish to ruffle his friend’s temper, …. and told that his advice was not wanted. But Corley’s brow was soon smooth again.

Similarly, this earlier sequence is rather more coherent, with the intermediate negation-carrying sentence (underlined), than it was before:

– First I used to go with girls, you know, said Corley, unbosoming; girls off the South Circular. as if he was conscious of being disbelieved.
But Lenehan could well believe it; he nodded gravely.

On the other hand, one should avoid becoming preoccupied with the putative coherence of the abridgement. It is not, after all, a version of the story that the reader ever encounters; it is simply a means of highlighting the story’s prospection resources.

The remaining prospection parameters can be searched for and added to the abridgement fairly straightforwardly. Searching for all narrative instances of mental processing and modal verbs (the final parameter) yields 14 distinct occurrences of verbs and projected clauses, 5 of which were already in the abridgement, and the 9 now added are mainly of quite minor importance, so that there is a strong sense of diminishing returns here.

As for frequent lexical keywords and clusters (shown in italics), the strategy adopted meant that that these, too, made only small additions to the abridgement. Only the keywords walk/ed/ing and street/s merited inclusion, and the extant abridgement already contained 11 of the 21 instances of walk/ed/ing and 9 of the 29 instances of street/s, so it was decided to forgo a separate search and inclusion of
the remaining instances, even though this meant excluding *He had walked the streets long enough with friends and with girls*, which comes in the middle of Lenehan’s FIT reflections. As noted earlier, there was only one cluster – *it off all right* – of any real prominence. The sentences containing all three instances of this cluster, and four of the five identified secondary sequences (ending with the first two words or beginning with the final two words of the cluster) had already been selected for the abridgement via one of the other parameters. The single exception that was now included was the sentence: – *She was... a bit of all right, he said regretfully.*

Parameter 6, highly prospective direct speech (and its inquit clause) was the final criterion to be searched for and included. There is of course no scope for an overlap between FIT text and prospecting direct speech; but on three occasions there are negation-sentences which are also prospecting direct speech sentences, e.g. Lenehan’s remarks *“I don’t want an introduction…. I’m not going to eat her.”*

The example just cited raises a general question about parameter 6. Does a negated direct speech prospection amount to no prospection at all? It seems to depend on the move type. In the case of questions and directives, such as Lenehan’s story-final *“Can’t you tell us?”*, the negation only weakens but does not deflect his prospecting request. On the other hand in negated declarative-structure in -forms and undertakings, the negation does seem fairly thoroughly to undercut the prospection. Thus when Lenehan says *I don’t want an introduction… I’m not going to eat her*, the negated prospections are quite attenuated ones (certainly by comparison with *I want an introduction... I’m going to eat her*); they carry weaker suggestions as to how things might actually proceed. But few of these prospective speech moves, even the negated ones, are of the minor and local variety noted above in Section 2, such as the *How much is a plate of peas?*, immediately answered by material not retained in the abridgement: *Three halfpence, sir.* Most are recurrent textual variants of the crucial event of ‘trying’ Corley’s girlfriend, and pulling off (all right) the taking of money from her.

Particularly because no independent search was undertaken for Parameter 2 sentences, those with a narrative-tense verb and a main character as participant, a question persisted as to whether a sufficiently large and representative selection of past tense narrative verbs are retained in the final 1,695-word abridgement. To try to resolve this doubt, a Wmatrix analysis of the abridgement itself was performed, and this showed that it retains 131 VVD items from the 285 in the original, that is, nearly half. Among the VVD verbs that have not been carried over into the abridgement are the following:

he rearranged the light waterproof …. He always stared straight before him….. ….They reached Stephen’s Green… they crossed the road… He turned to the left when he came to the corner … felt more at ease… suited his mood… He paused at last… He eyed this food...he pushed his cap back… He ate his food…he made
a note... This vision made him feel... He went into Capel St...he turned into Dame Street...he met two friends... he lit his last cigarette... A woman came running... His broad figure hid hers from view. He took them as a warning... he ran eagerly across the road...

As far as prospection and core narrativity is concerned, I would submit that the vast majority of the predicates listed above are appropriately absent from the narrativity abridgement. Some have a strictly local relevance (he rearranged the light waterproof; he lit his last cigarette), several relate to Corley and the young woman but only supplement what has already found its way into the abridgements; and a considerable number report Lenehan's following of the couple, and then his own wanderings through the city. Those relating to his frugal meal are again only an enlargement on those sentences reporting this episode which have already been captured. If the analyst uses the VVD category in searches, it is important to bear in mind that this Wmatrix category excludes all instances of was (he was a sporting vagrant) and past tense had (she had her Sunday finery on; He had a brave manner of coming up to a party), on the grounds that these are not lexical (but other intensives like seemed and became are included). It also excludes all passives (a yachting cap was shoved far back from his forehead), and perceptual-intensive looked as in his face looked older. Thus one needs to bear in mind that not all past tense verbs fit the VVD frame (although nearly all dynamic or action ones do).

The narrativity abridgement is not presented as the definitive extraction of that material most centrally involved in narrative progression in “Two Gallants”, but as a work-in-progress best approximation. My own first reaction is that the abridgement is too long, i.e., that it has not sufficiently constrained the search and identification of high narrativity sentences. How might further constraints, and further reduction of the abridgement, be applied? One strategy might be to require that sentences only qualify for inclusion in the abridgement if they contain instantiations of at least two (or even three) parameters. A rather different procedure would be to present the abridgement to readers and ask them to highlight those sentences which they feel contribute most (or least) to their impressions of the story's development or dynamism. It is possible that informants will tend to agree in finding sentences from one parameter insignificant (and those from another parameter highly significant) to narrative dynamism: good grounds for removing that parameter from the model (or promoting the other in it). Only testing will clarify these matters.

The eight parameters of the narrativity/prospection model are also open to internal adjustment in various ways. Two such adjustments I will simply note here. One adjustment would be to remove those sentences which figure here only on the grounds that they contain the second-most frequent keyword Lenehan; unlike the
Corley sentences, these seem to be of lesser importance to the creation of narrativity. We earlier saw that sentences containing the topmost keyword seem to behave distinctively; but the fact that Lenehan enters into long cohesive chains (via both proper name and pronominal repetition) was an argument for the inclusion of at least those sentences in which he, like Corley, is proper named. In practice that enlargement of the text is relatively slight: about 150 words (while Corley-containing sentences amount to 950 words of the text). But the slightness of the expansion may equally speak to these sentences’ secondary role. If the inclusion of the Lenehan sentences can be questioned, the exclusion from the abridgement of those sentences referring to the third character, the woman who spends the evening with Corley, arguably becomes less problematic. No direct or indirect naming of her surfaces as a keyword in the text; and indeed the pronoun she is a negative keyword (disproportionately infrequent) in the text, occurring only 21 times.

A second adjustment relating to the parameters would discard all those sentences that are included here only because they contain a bound negative morpheme (e.g. noiselessly, disordered), retaining only those sentences where a free negative morpheme (not, nothing, etc.) is used. One could go even further and require that such negations qualified the predicate and not some other phrase (cf. Not far from the porch of the club). These two adjustments, capturing only the sentences with the topmost character-naming keyword and discarding sentences with only secondary negations, would reduce the abridgement by a further 16% to 1,321 words, i.e., to about a third of the original story. But they are only two among numerous adjustments and improvements that, over time and after more testing, may be advisable.

These two chapters have proposed and described a matrix of eight important parameters of narrative prospection in short stories. They have tried to demonstrate that semi-automatic procedures of text-extraction can highlight the main threads of prospection in a story. In principle these parameters are central to progression and reader expectation in every kind of written narrative, from the one-page folktale to the Victorian novel. But the present focus is the modern short story. The theory is that readers especially form expectations as to what will happen in the subsequent narration from the guidance of particular parts of the earlier narration. The first sentences of each narrative paragraph, I have argued, are especially indicative. So too are sentences in which a frequently-named character is mentioned; as are ones in which such characters act or are acted upon, or their directly reported proposals or expectations about the future or their free indirect thoughts are disclosed; as are those in which high-frequency lexical keywords or prominent negation or projecting mental or modal verbs occur. These eight streams or threads are (I propose) especially important for narrative progression.
Much still remains provisional and at a preliminary stage, needing further analysis and testing. I cannot be certain these eight parameters, as here characterised, are incontrovertibly and solely the core narrativity resources: such questions are not amenable to straightforward scientific proof, although testing and accumulating evidence can strengthen – or rebut – the model. Are all eight equally important to narrativity? Almost certainly not, but again, the proposed model is at far too early a stage for the making of confident predictions about parameter-weighting. Are the parameters a coherent or unified set, a system, governed by just a few fundamental principles such as agency, mentality, and repetition? In a sense, the heterogeneity of the elements is disappointing to me, as it may be to others, since ideally one hopes to uncover a closed and homogeneous code or grammar, whereas the present set more resembles a bundling of disparate devices. But such variegated complexity perhaps best answers to our sense of the polyphonic nature of narrative construction.

Now my hope is that it might be possible to show on independent grounds (e.g., in readers’ responses) that the parameters especially contribute to a story’s narrativity, i.e. a reader’s sense of the narrative’s progression. The devilish beauty is in the detail, however: the next chapter explores the usefulness of this foregrounded network of prospection in the creation of narrative suspense and surprise.
1. Sentences in which occur the top-keyword naming of a main character.

2. Sentences containing occurrence of narrative-tense finite action/dynamic verbs, where a (frequent keyword) main character is an argument or transitivity participant (e.g., Subject or Object; or Actor or Goal).

3. The first sentence of each narrative paragraph or section.

4. Sentences containing “fully lexical” frequent keywords and clusters (where “frequent” keywords are defined as those which comprise at least 0.10% of the text, or that occur at least 5 times in the text).

5. Sentences containing characters’ represented thought (especially FIT and DT). Three ordered procedures for automatic identification of FIT:
   i. select all non-direct speech sentences where a modal verb (would, could, should, might, etc.) occurs with a directly adjacent personal subject pronoun, except where the pronoun precedes the modal and not/n’t follows, or where the pronoun+modal is preceded by a subordinator such as if, whether, because, although, or similar.
   ii. include all questions and exclamations in the narrative. all sentences in the narrative with a narrative-tense verb and ending with a question mark will identify all interrogatives. For exclamatives the search should cover two patterns: all narrative (non-speech) sentences ending with an exclamation mark; and those with an initial Wh-word, and a narrative tense verb – often BE – that is not immediately followed by a, the, or some.
   iii. include all flanking sentences containing a subjective modal, such as could, would, should, must, might, surely, certainly, probably, possibly, and maybe.


7. Negation-carrying sentences: not, n’t, never, no, nothing, nowhere, fail (to), -less, without, dis-, un-, im-, in-, etc.

8. Sentences containing narrative verbs of modality and mental processing, especially where these are followed by a full narrative proposition as complement. The most important such verbs include: know, think, seem, appear, feel, suspect, expect, want, need, see, look, wonder, believe, and realize.

**Figure 4.** Summary of the eight proposed parameters (and search procedures) of narrative prospection.
The textual tracking of suspense and surprise

This chapter is about the elements of narrative suspense and surprise, and the degree to which their occurrence can be identified by examining the particular configurations of phrase and sequence in the foregrounded and abridged version of a story. While keeping in mind that these cueing sentences are not read in a vacuum but against a contextual background of knowledge and values, I here ask whether we can see suspense and surprise being generated by means of the eight narrativity parameters postulated and detailed in Chapters 7 and 8.

1. Narrative suspense

Narrative suspense involves imposing on the reader. What is imposed is noticeably-delayed negative-outcome foreseeability. These conditions will be elaborated upon below, but first consider the following simple case. A character flies from Paris to Toulouse; upon her arrival there, anything can happen: such general uncertainty is not suspenseful. An extremist embarks on a flight from Paris to Toulouse, intent on blowing up the plane; again anything can happen, but we particularly foresee that the extremist may succeed in blowing up the plane, or she may fail. This is potentially suspenseful. If in the telling the reader or film-viewer judges that there is a noticeable delaying of the narration of the outcome or prolonging of the narration of the intervening material, then suspense is achieved. Among other means, suspenseful delay can be achieved by a narrative pause in the course of telling inherently dynamic unfolding events.

Narrative suspense differs from kinds of merely grammatical and discoursal suspense, which are primarily a matter of perceived incompleteness: a grammatically or discoursally required completion is missing – postponed (briefly or permanently) beyond its expected place of occurrence. Consider the following instances of incompleteness:

1. He put the book on the …
2. He placed on the shelf …
3. She ate…
4. Have you seen my car keys? …
5. Two capuccinos, please.
In (1) and (2) the reader awaits the noun and the noun phrase, respectively, required to complete each sentence. In (3), *She ate*… can be continued, grammatically, by all manner of semantically-diverse phrases or clauses, including none. By contrast *He put the book on the* … raises the strong expectation that a construction-completing noun will occur very soon if not at once, and we expect that noun to be a notionally flat horizontal surface: *newly-mown grass; floor; roof of the car;* etc. Examples (4) and (5) are simply reminders that speech exchanges in typical contexts of use can generate local discoursal suspense, where a question prospects an answer, or a request prospects a compliant act or a justified declination, and where these answering moves are noticeably delayed. To these could be added the periodic sentence of Milton and others, or examples of ‘arrested’ sentences such as the following from Joyce’s *Ulysses*:

Looking down he [Bloom] saw flapping strongly, wheeling between the quaywalls, gulls.

But narrative suspense, which manipulates story, goes further than the examples just discussed, which only exploit sentence or discourse structure. Narrative suspense is a function of the reader’s strong, deferred and conflicted expectations. They are conflicted because readers foresee a likelier dispreferred outcome and a less likely preferred one. They are impatient for the outcome, but not if it is the more likely one. Plot wish-fulfilment primes the reader for one outcome, but death (or one of its transformational derivatives: sickness, failure, loss, breakdown, blockage, invisibility, inaudibility, blackness, silence) informs the counter-expectation. We have a particular expectation of that which we particularly abhor. So fear is also involved in the most powerful instances of suspense. And our readerly experience of these conflicted expectations is rendered more acute by prolongation of the uncertainty, delayed telling of the resolution which is effected by a slowing of the pace of telling to scenic narration or even descriptive pause, or by digression, or by ‘cutaway’ to a loosely contemporaneous related plot. By contrast with the delayed occurrence of a main clause or an obligatory clause-element in a sentence, where foreseeability is chiefly structural and stylistic effects tend to be local, narrative suspense is considerably more experiential, and emotionally as well as intellectually engaging.

Furthermore, the reader’s strong foreseeing of both a good and a bad resolution only makes the delay more palpable, even an affront. We become more conscious of the narrative’s constructedness, its design and composition by *someone* (author or narrator) of whom we may come in the interim to be critical and suspicious: why have they not told us by now what we can so clearly predict? What are their motives? Are they teasing us, mocking us, insincere and dissembling, disingenuous, or just obtuse and incompetent? These strong reader reactions may be
why Barthes, in his famous introduction to narrative analysis, called suspense “a veritable ‘thrilling’ of intelligibility” (1977: 119).

The necessary textual conditions for narrative suspense (briefly sketched in Toolan 2001; see also Iwata forthc. for a very detailed analysis of the narratological conditions for suspense and surprise) can be grouped into three:

a. **Forking foreseeable outcomes**
   The narrative ‘forks’ in a Barthesian sense of reaching a point of development where very few (often just two) alternative continuations or outcomes is highly predictable, so that two (or just a few) narrative completions are clearly ‘foreseen’ by the reader (e.g. survival vs. death, being caught vs. not being caught, someone accepting or rejecting a proposal, a bid succeeding or not). All foreseen outcomes are predictable because the narrative is usually here invoking sequences that are schema- or script-like in their recognisability to the reader. That script-like foreseeability is one of the ways suspense can be distinguished from less sharply definable passages of ‘tension’ or ‘anxiety’. For example, the protagonist is driving his car at high speed in treacherous conditions and suddenly encounters black ice or an obstacle in the road, as in the near car-crash in McEwan’s *A Child in Time*; or an agent on a secret mission, trying to smuggle vital documents out of the enemy country, is approaching the border, where guards will inspect his papers; or someone is trying to decide whether or not to confront another protagonist with some potentially life-changing information. In each case, the script-like familiarity of the situation makes a restricted set of outcomes highly foreseeable: the speeding car will crash, or it will narrowly avoid crashing; the agent will be caught or shot, or he will escape. The outcomes are often weighted, by our real-world knowledge and contextual evidence, as a more likely and a less likely one.

b. **Someone cares greatly about the outcome**
   The second condition reflects the evaluation or emotional involvement that many theorists (e.g. Labov, 1972; Carroll, 1996) have emphasised arises in suspense. The reader is aware of a preference (moral or ethical, or otherwise) for one outcome over the other or others. That preference may be rooted only in a character, but ideally readers are sufficiently engaged with the characters and the story world that they much prefer one outcome as happier, or more fitting or interesting – even though, typically, the abhorred or undesired outcome seems more probable and realistic in all the circumstances. These outcomes are strongly preferred and dispreferred, respectively, because they relate to issues that are highly relevant to the character (and often, vicariously or empathetically, to the reader), and highly consequential. The reader’s awareness of suspense is more important than that of any protagonist: in some narratives, the reader will see and feel suspense where the protagonist is blithely unaware of imminent crisis. But it is possible, through limited moral and emotional empathy with the relevant character, for the reader
to feel little in the way of anxious suspense when processing a progression which a character evidently finds suspenseful, as at the end of “Two Gallants”.

c. The telling delays us
At and after the point of perceived narrative forking among broadly predictable completions, all such completions are for a time felt to be ‘withheld’: the disclosure of just which completion obtains is noticeably delayed, beyond its earliest reasonable report. This perceptibly delayed narrating is itself, in Gricean terms (under Quantity and Manner maxims), implicative of difficulty, a problem, suggesting to the reader or viewer that the less desirable outcome may indeed be about to transpire. The sense of delay and withholding is achieved by a local expansion of narrative duration, thus of ‘taking longer’ to move from an awareness of the coming crisis to its outcome, or by a diversion to discussion of tangential matters, or an ‘expanding’ of the moment into the psychological time of the anxious protagonist, or by some other means. In written narratives, the language of the delaying text signals by omission and commission that it is suspense-generating: the major omission is the not telling at once of the outcome.

A simple example of Condition (a) would be where a character in a story decides to smuggle an illegal item through customs. Depending on the nature and intended use of the illegal item, your empathy with or antipathy towards the smuggler, the apparent efficiency and decency of the customs officers, and similar factors, the reader might strongly hope that the smuggler gets caught but judge it fairly unlikely that the person will (b). Several pages of narrative might then report the character’s journey to their point of departure from the first country. Provided that the narrative is no slower or more digressive than previously, then none of these pages would be suspenseful in the sense intended here, although they could be ‘gripping’ and anxiety-inducing. But once the character reaches customs (c), so that imminent reporting can be expected of whether or not they are stopped and caught, then any delay or elaboration of those next steps creates and constitutes suspense.

Where it occurs, suspenseful text is likely to be characterised by material that is anything but provision of a report, in the narrative tense, of the anxiously-awaited outcome: descriptions of states, background causes, moods, settings, ongoing circumstances without clear temporal limits, analeptic or proleptic narration (that is, ‘flashbacks’ and ‘flashforwards’), hypothetical and imagined outcomes, and parallel contemporary events. In short oral narratives of personal experience, the general term for all these enrichments and deflections of the bare ‘what happened next’ is evaluation (Labov 1972); but that term is too general for the array of suspense techniques deployed in film and literary narrative art.

The three suspense conditions above do not include any requirement to the effect that a resolution or outcome must be disclosed, so as to bring an end to the
prolonged uncertainty. In Raymond Carver’s very short story “Little Things”, a couple on the point of breaking up get into a tugging match over which of them will take the baby, and the story ends at the suspenseful high point, with no clear indication as to which of the couple ‘wins’ the baby, or whether the baby is harmed or killed in the process. So the story ends on a note of suspense, a note which reverberates indefinitely. Arguably the suspense only begins at the very close of the story, when we discover that we are not going to be told the outcome. In this respect it differs from another Carver story, “The Bath”, in which Scotty is knocked down by a car, and then lies for several days in a coma at the hospital, attended by his desperately anxious parents. From soon after Scotty’s removal to hospital, then, the story is suspenseful as we, like the parents, wait for a complete resolution. But the story ends with the opening, only, of a phone call (“This has to do with Scotty, yes…”) to the boy’s mother, who has gone home to shower and get a change of clothes. So again, no end to the suspense is reached. That suspense has begun well before the story’s close, however. Carver published a different version of the same initial narrative, with the title “A Small, Good Thing”, but that narrative takes events well beyond the cryptic phone call that closes “The Bath”. This longer version is discussed below in Section 4.

It should also be conceded that if a fuller mapping out of kinds of suspense is undertaken, the simple binary contrasts invoked in this chapter may need replacing by a model that is more scalar. For example, I have described resolution as if it were required for a story to bring suspense, a delayed reporting of outcome, to an end. But in response to some short stories readers are inclined to say that, having created some suspense, the story progresses to an unresolved close without a strong persisting sense of suspense. A master contemporary exponent of this technique, in which a suspenseful question gradually bleaches away without there being any simple-minded undertaking to narrate ‘the final outcome’, thus without resolution, is Colm Tóibín (e.g., several stories in *Mother and Sons*).

Further examples abound, in highly crafted literary stories of the kind that have sometimes been assumed to be too subtle to exploit such vulgar reader-manipulative devices as suspense and surprise. In Kipling’s “Mary Postgate”, for example, suspense is crucial. Mary finds an enemy airman in the garden, shot down and unable to move (the story is set during the First World War): will she give him humanitarian aid or not; or worse, might she even act to hasten his death? At the end of Joyce’s “Eveline”, in the course of which some suspense has been built up – will she leave with Frank or won’t she? – most readers judge that the suspense is released, as Eveline stays. Mild suspense arises in “The Dead” as an atypically-aroused Gabriel follows Gretta up the stairs towards their hotel room, only for him to suffer the reversal of surprised discovery that “what Gretta is thinking of” is
nothing conjugal and imminent, but Michael Furey long ago, who may have died for love of her.

Consider Katherine Mansfield’s short story, “Bliss”. The story thematises Bertha’s ‘bliss’, her sense of charged happiness at being who she is and where she is, and now, at the dinner party that is the story’s backcloth, perhaps for the first time alive with sexual desire for her husband Harry. In her own constrained way, she is impatient for the guests to go. She is a coiled spring of happiness from the story’s opening words, but the situation is not one of narrative suspense, for reader or, except mildly in the final stages of the dinner party, even for Bertha. But then, to her and our devastating surprise, at the very close of the evening, Bertha discovers that Harry is passionately involved with Miss Pearl Fulton. The textualisation of this surprise will be examined in detail later in this chapter. By cruel irony, Pearl is the same cool enigmatic beauty with whom Bertha has only recently thought she had achieved a special communion, when they jointly admired a pear tree in the garden. Turning again towards that aroused and glittering pear tree, Bertha ends the story on a note of painful suspense: ‘Oh what is going to happen now?’ Thus, painful surprise is fast followed by unreleased suspense.

What of suspense in crime and detection stories? It is a requirement of the genre that identification of the true perpetrator be palpably delayed, and this creates a kind of suspense. But such suspense over identification is different from and less powerful than suspense with regard to plot-development; along the way to the belated identification, it is quite common for a series of surprises, including further murders, to occur. Besides, crime stories are rarely as narrowly framed as is classic suspense: few crime stories inform us at an early stage that just one of two characters is the murderer. And there is a different role for foreseeability: we do not foresee what happened to the (first) victim; on the contrary we are usually told this at an early stage. What we need to puzzle out, with guesses rather than foreseeability, is who did what happened.

In this section a number of claims have been made about what the reader or viewer infers and prospects, along the route to experiencing suspense. To some researchers interested in discourse processing, such as cognitive psychologists, some of these claims would be highly questionable and even psychologically invalid. In the psychology research community, it is a matter of considerable debate just how much inference-generation goes on during online (real time) processing and comprehension of narratives (as distinct from post-hoc formulation of causes and explanations for event sequences). Several theories have been proposed, but an interestingly extreme view is the ‘minimalist position’ of McKoon and Ratcliff (1992), which argues that readers make very few online inferences during narrative processing. Thus they suggest that, for example, the sentence ‘The actress fell from the fourteenth storey’ does not prompt the inference that death will follow,
unless its generation is needed to meet ‘local coherence’ requirements. But for the
narrative stylistician those ‘local coherence’ requirements need detailed attention.
Within what genre does the actress sentence occur? Does the actress have any
special powers? These and many other questions are prompted by the example,
which suffers from the familiar limitation of being too decontextualised.

At the same time, we do need to constrain the postulating of inferences that a
given text is claimed ‘inevitably’ to trigger in the reader. Corpus-based narrative
stylistics is most useful when it attends only to those inferences and projected out-
comes whose sources can be identified in the text; in other words, it should attend
as much as possible to just those inferences, reactions, and prospections that de-
monstrably have a textual source. The stylistician privileges the textually-instanti-
ated over what may or may not be said about narrative suspense in more abstract
and generalised terms. Narrative suspense is in the text, I contend, and is so en-
trenched that it often persists though multiple readings. That is, written narrative
worse is not necessarily neutralised for readers after a first complete reading:
contrary to some theoretical predictions, readers often find a story almost as sus-
penseful in a second or third reading as first time around (Hoeken and van Vliet
2000; see also the evidence reported in van Peer 2007c). This suggests that the sus-
pense arises not from the not knowing what will happen (the second-time reader
does know) but, given all the circumstances, from the not being promptly told.

2. Narrative surprise

Narrative surprise is created in the reader where the following conditions, some-
what in contrast with those for suspense, combine:

(a) the narrative has approached and may be presumed to be passing an un-
problematic ‘milestone’, where there is little or no sense of potential forking into
different sequels and where, rather, a stereotypical or schematic next event or scene
is strongly predicted. There is a well-primed script in place, we think. In Heming-
way’s “Indian Camp”, for example, the stereotypical story schema is in part that of
an anxious husband witnessing his wife’s difficult labour, one requiring an emer-
gency Caesarian operation with crude and limited equipment. Since all goes well
and the associated suspense is released with mother and child emerging safe and
sound from the ordeal, one expects that when the narrative focus turns back to the
father, it will be to describe his relief and joy, and gratitude to the doctor, the mini-
god controller of good outcomes.

(b) the expected ‘non-forking’ schematic/automatic progression does not go
through as predicted, and something unpredicted and seemingly unprospected
happens, something which is relevant and imaginable in the circumstances if only
we had operated with full foresight or insight. This foreseeable but unforeseen development pulls us up short, and may cause us to re-assess much of the narrative whose shape and direction we thought we already understood. In fact, the relations between surprise, accident, and coincidence deserve a full study of their own.

The idea that in narrative surprise ‘something unexpected happens’ needs to be clarified: what ‘happens’ at the point of surprise is an act of narrative disclosure. While this typically becomes part (the closing part) of the surprise sequence for the reader and one or more characters, the surprising conditions usually reach back in time to an earlier unnarrated point in the story. In “Indian Camp” the surprise is felt when the Indian husband is approached by Nick’s father after the successful Caesarian delivery and is found to be dead, having killed himself evidently to escape the torture of his wife’s screams of pain. The key surprising event, the act of suicide, obviously predates its belated discovery and narration. But even if surprise is in such cases a ‘telling late’, it is never analeptic: there is no sense of the narration going back to tell what happened earlier (even if a character figuratively ‘goes back’ in time, as Gretta does in her surprising revelations about herself and Michael Furey, in “The Dead”).

In narrative surprise a reader experiences a new development as unforeseen but upon reflection foreseeable. To say that the event is, on analytical reflection, foreseeable is also to say that it ‘fits’ the larger structure of narrative conditions and developments in the entire story: it is not a detached addendum, but a fact or outcome that can be fully integrated with everything else, filling a gap we had not even noticed was there. In the same vein, David Lodge in *The Art of Fiction* (1992) prescribes that tellers must provide the reader with enough information so that the belated revelation, when it does occur, is both unexpected but also convincing (the author must tell neither too little nor too much). Surprises are most effective when they are felt to be in no way absurd or inexplicable, but reasonable, possible, and even probable. The more probable we belatedly judge the surprise to be, the more we are chastened or humbled by our own corrected and better judgement.

To experience a really effective narrative surprise is to be caught up in an activity of self-teaching, of reflexive critique, and this is part of what makes them so valuable. They do not entail simply a revision, by the reader, of their grasp of the narrative; with a little jolt of correction, they also compel the reader to examine and find lacking their own understanding. So these surprises are finally about understanding, not merely information. In ordinary circumstances, no shame or chastening accrues to the reader who, before being so informed, fails to foresee that a particular character has brown eyes, or was born in Buenos Aires, or checked out of the Savoy hotel at 10 am on the Monday morning; it is entirely reasonable not to have foreseen such particular new narrative facts prior to their telling. By contrast genuine narrative surprises present us not merely with unforeseen
information; they display for our contemplation a moment or a space where our understanding (of facts and events integrated with motives and psychology) has failed. In doing so they confirm the claim made in Chapter 3 of Proust’s *Time Regained* (1933: 322): “Every reader, as he reads, is actually the reader of himself. The writer’s work is only a kind of optical instrument he provides the reader so he can discern what he might never have seen in himself without this book.”

There are ways in which suspense and surprise are partners, and even sometimes in a kind of complementary distribution, where a reader caught up in a suspenseful situation is struck, at the very point of its releasing resolution, with a distinct and unforeseen surprise. Conversely a surprise development can lead to a suspenseful passage in which a protagonist succeeds or fails in the changed circumstances. Narrative surprises seem typically to involve a downturn in a character’s fortunes, plain sailing abruptly become shipwreck. When narrative suspense is resolved, it is usually resolved ‘happily’, a triumph of hope and subjectivity over everyday likelihood. Taken together these ‘ups and downs’ of a protagonist’s progress may account for the fact that ordinary readers often use the ‘rollercoaster’ metaphor to describe novels filled with suspense and surprise alternations. But there are also ways in which suspense and surprise differ: the former is felt in the course of a (protracted) episode and ceases with its outcome; the latter is punctual rather than durational, felt acutely in the unforeseen outcome, while its message about our fallible understanding endures.

It follows from the foregoing conceptions of narrative suspense and surprise that they need to be distinguished from other important text-shaped phenomena experienced by readers in the course of narrative processing, such as tension, secrecy, mystery, foreboding, and inevitability. Some of these are more precisely definable than others, and all involve manipulation of what is told and the time of telling: they all exploit what has been called the double logic of narratives, the fact that a story tells in time what happened in time but need not follow, in the order and amplitude of the telling, the chronological order and apparent importance of the events that it relates. Thus, for instance, a seemingly major event that occurred early in the chain of events may be reported relatively briefly, and belatedly, in the narrative discourse.

It is worth remarking also that narrative suspense and surprise are closely bound up with the crucial business of narrative closure. Where suspense is prominent in a story, it tends to figure especially at the end, and its completion or otherwise is often coterminous with the end of the discourse: it is much rarer to find major suspense raised and resolved early in a story.

A final point we can note is that suspense and surprise are two of the most basic means by which a crafted narrative plays havoc with orthodox notions of a narrative arc or trajectory, and metaphors of a smooth, arrow-like flight of plot.
from the teller’s bow to a target (story-completion). With suspenses and surprises, it as if the arrow or trajectory abruptly changes course or speed for a period, not in accordance with the laws of Newtonian physics but in outrageous defiance of them, before perhaps resuming the earlier course and speed. And these breaches of smooth predictability are not random or inexplicable, but carefully designed and motivated.

3. Textualising suspense in “Two Gallants”

Given the three-element characterisation of narrative suspense proposed in Section 1 above (a plot-important forking, a reader or character preference about outcome, and a delay), it is possible to describe the general textual features that are likely to arise in the representation of each of those elements. Thereafter those textual expectations concerning each of the elements can be matched with the actual textual evidence and, in the present case, with what is here claimed to be the foregrounded role of the material in the core narrativity abridgement.

The first element, an important forking in the plot, with both (or all) outcomes clearly foreseeable and even describable as imagined outcomes, will unsurprisingly tend to be reflected in the simplest cases by conjunctions associated with alternatives, such as or, whether, and if clauses. We can also expect future-oriented modal constructions using might, could, should, would and so on, in declarative mood or in polar interrogatives: X would happen; then [or time adverbial with span reaching into the future: in two days, in two years] x would happen; Would x (happen) (or not)? These will often implicitly express a character’s point of view, which of course is explicit where these constructions are within the direct speech or thought. In DS and DT, besides might, could, etc., verbs like will, can and be going to, and imperative sentences, are also options. Often, projecting mental processing verbs which at least partly paraphrase the idea of foreseeability will be used: expect, foresee, know, wonder, fear, hope, etc. Other modal and epistemic means (verbs, adjuncts, and more dispersed lexical constructions) expressing the uncertainty of the outcome will often figure in the suspenseful text, as auxiliary elements. The script-like foreseeability of the outcomes is reflected in the fact that the dispreferred but expected one of these is sometimes alluded to through indicative noun phrases which mention predictable phases or components of that outcome. Thus during a period of suspense over whether a loved character will live or die, there might be phrases like the death-bed scene, the crowded funeral ceremony, the freezing cemetery, the gloomy reception, the Will, and the family feuding. When used with care, simple search engines on the internet (‘googling’) can powerfully confirm
what words and actions over and over again have been part of a familiar script or schema for ‘tragic premature death of loved one’.

Are such elements apparent in the latter half of “Two Gallants” – where readers have remarked on a building of suspense – and in particular in the prospection abridgement, presented in Section 5 of Chapter 8? They are, in abundance. To begin with, the abridgement records Lenehan’s reiterated

Well... tell me, Corley, I suppose you’ll be able to pull it off all right, eh?

and

are you sure you can bring it off all right?

together with Corley’s testy reply:

– I’ll pull it off, he said. Leave it to me, can’t you?

As noted earlier, it off all right is the only thrice-occurring four-word cluster in the story; on this first encounter with it the reader cannot be aware of its whole-text significance, but then within a few lines it is prominently repeated. These prominent repetitions thus initiate a crucial network of mentions of bringing something off all right (as discussed in s.3 of Chapter 3 and s.4 of Chapter 7). Corley’s irritation indicates that the question is not entirely trivial. Similarly reinforcing is Lenehan’s parting “Work it all right now”, three words of which echo the it off all right cluster. As a small taste of the later suspense Lenehan suffers, Corley, we are told, “did not answer”.

Thereafter, abridgement sentences like The problem of how he could pass the hours till he met Corley again troubled him a little confirm that for Lenehan meeting up again with Corley is or is linked to his and hence the story’s main goal. The reader does not know what the it is that Corley is to bring off, but can already see that it is important to Lenehan. The precise referent of the it remains unclear until the very end of the story, when Corley displays the gold coin he has obtained. The pronoun is therefore a good instance of the partial ellipsis that Prince (1988) has called the ‘nonnarrated’ (entities whose full identification is delayed, or even left incomplete indefinitely); and nonnarrateds are always potentially suspenseful.

While building suspense over whether Corley does pull it off, the prospection abridgement also introduces a secondary suspenseful anxiety of Lenehan’s, namely that Corley might not keep their rendez-vous at all:

[He] kept his gaze fixed on the part from which he expected to see Corley and the young woman return.

The abridgement textualises again the bifurcating outcome:

He wondered had Corley managed it successfully
and Lenehan’s preferences:

the memory of Corley’s slowly revolving head calmed him somewhat: he was sure Corley would pull it off all right.

So again the preferred outcome is textualised, in irrealis mode, that is, as a possibility and not a reported actuality. It is this repetition of ‘event release’ outcomes in the irrealis mode, in the textual space after a prospection or ‘event onset’ where realis completion is reasonably expectable and would be timely, that constructs suspense.

But Lenehan is “All at once” entirely unsure:

perhaps Corley had seen her home by another way and given him the slip

and this leads to the question

Would Corley do a thing like that?

Soon that polar interrogative is being answered in the pessimistic affirmative:

They must have gone home by another way.

No sooner has this imagined failure been echoed by the breaking of his last cigarette than Lenehan’s supposition is contradicted: he sees the couple coming towards him. Soon after, again, negative prediction resumes: He knew Corley would fail; he knew it was no go. But there are more than 300 words of further text in the story, including Lenehan’s reiterated demands that Corley narrate the outcome ( – Well? he said. Did it come off? and – Can’t you tell us? he said. Did you try her?) before we reach the final report that a now-smiling Corley is the bearer of a small gold coin. All the way to this conclusion, the textual signalings of the outcome, nearly all of which are caught by the prospection abridgement, have been muted or misleading. Included among them is the antepenultimate sentence, Corley halted at the first lamp and stared grimly before him. In that sentence, the ‘staring grimly’ is likely to be interpreted as indicative of Corley’s failure; there is a semantic-kinesic mismatch here: people who pull something off all right do not stare grimly.

The second noted element of narrative suspense is the preference of the character or of the reader (often of both) for one less likely outcome over other more likely ones. The reader may tend to prefer whichever outcome most fully meets the main lack identified in the story’s beginning Situation. In “Two Gallants” it is only Lenehan who has a clear and strong preference about outcome, and not the readers: they do not know precisely what the ‘it’ is that Lenehan hopes Corley will pull off, and might well not empathise with Lenehan if they did. Textually, outcome preference may be reflected in reports of the character’s thoughts or words, particularly their wishing, hoping, etc (boulomaic modality) and even the more
borderline epistemic/boulomaic **wonderING** if/whether. These preferences may be reported as general ones in the situation: *everyone was hoping that A*. Preferred outcome is also sometimes inferable via reporting of an antipathetic character’s views: *only X said/thought/hoped that not-A*. But the most pervasive signaling will be in and through the lexis, and the associations and prosodies or primings of that lexis. And here it seems we need to conceive of lexis increasingly in terms of what Sinclair has called the lexical item (Sinclair 2004: 121–148), i.e., not single words (such as *bring* and *off*) evaluated separately, but sequences of words forming lexical items or phrases. Thus to *bring or pull something off* is normally to achieve a difficult but valued task or goal; and (in non-ironic uses) *to fail* is always a bad thing, and to *manage [something] successfully* always a good thing, so attributing these to Lenehan (*He knew Corley would fail* etc.) makes clear his preference.

The final crucial element is delayed disclosure of outcome. Often prominent in its textualisation is greater frequency of temporal expressions of all kinds, of which the following are only a selection: *meanwhile, for x seconds/minutes/hours, eventually, finally, at this point, before this/that*. There may also be greater use of (scenic) direct discourse, or, where the latter is already the norm, noticeable local repetition of words or phrases, which are an enacting of prolongation (a breach of the Gricean manner maxim, which enjoins succinct orderliness). Delayed disclosure exploits the duration or pace characteristic of narration, often creating paradoxical effects, such as a story emphasis on hurry and haste, but a slowness in the discoursal reporting of significant developments. More important than any of the above, as explicit marker of delay, is the relative extent of ‘intervening’ text: text which does anything but report the completive outcome that has been clearly prospected, and is perceived by the reader as falling between prospection and outcome and therefore delaying that outcome.

Again these temporal characteristics are extensively apparent in the final third of “Two Gallants”, from the point at which Lenehan, waiting as agreed at the corner of Merrion St and thinking it is *surely half-an-hour since he had seen the clock of the College of Surgeons* on the stroke of ten, expects to meet Corley and find out whether he has succeeded. More significantly yet, many of these temporal features are inherited by the abridgement (but not all of them, it should be conceded). Among temporal expressions retained are the following: *All at once, half-an-hour, Suddenly, quickly, quick, while, at once, remained, swiftly, hurried, then, continued, as before, Again, had reached, Still, and kept up*.

In all the sentences in which the narrative is telling us of Lenehan’s uncertainties and speculative foreseeings (suspense elements a and b), the text is not telling
what actually happened. It is important that the telling, at last, of the couple's arrival should be prefaced by the time-marker Suddenly:

Suddenly he saw them coming towards him.

This hints at the idea, counter to our impressions, that Lenehan has been inattentive in his waiting and looking; it also seems to promise a rapid advance in the action, and imminent disclosure of the outcome. But any such expectation is deflected. We are told what Lenehan sees – the couple's walking quickly, their uncompanionable silence – and what he concludes from this evidence: He knew Corley would fail; he knew it was no go.

Even in the abridgement of the story, the temporal markers of hurry and haste without resolution are retained: Lenehan follows them at once, Corley remained standing and then when the door closes on the woman he began to walk swiftly. Lenehan then hurried on in the same direction, and when he calls out Corley continued walking as before. From the he cried again in the abridgement we can deduce that Corley is deflecting his friend's obvious requests, three times, for the good or bad news. Thus Lenehan's enquiry is repeated several times, and this should be set beside other repetitions such as He knew Corley would fail; he knew it was no go. All such prospecting repetitions create delay; they also allow us to infer a continued absence of completive conclusion – in the same way that when someone says “Pass the salt please… pass the salt please” we infer that the addressee has not complied immediately after the first request.

When significant cues of all three suspense elements appear even in the story's prospection or narrativity abridgement, then this is good confirmation that suspense is present.

A preliminary attempt at mapping the textual creation of suspense, then, might simply count the number of words or sentences that fall between a second textual prospection, which might reasonably be expected to be shortly followed by a completion (the outcome), and the point (if any) of actual textual completion; where no completion is textualised, this amounts to permanent postponement. What is striking about “Two Gallants” is that because its latter half is so extensively narrated from Lenehan's limited viewpoint, the text can create premature (but at the time of reading, valid) expectations about completion of the prospection. Thus when the reader encounters Lenehan's despairing sentence They must have gone home by another way they might suppose that this is a valid completion – only to find this supposition corrected two sentences later: Suddenly he saw them coming towards him. So they must have gone home by another way is a first reasonable completion, even though its status subsequently changes. From this point to the display of the gold coin there runs a 419-word suspension.
As already stipulated, the prospection must, at least implicitly, be of major importance in the story, and tend to engage the preferences of one or more characters and/or the reader’s empathy. These will tend to be textually manifest, through repetitive mention of one or more foreseen outcomes, and through character- or reader-aligned modality, mental processing, and other evaluative means: wishing and wanting and fearing and anxiety and uncertainty. We have seen, confirmed in the abridgement of “Two Gallants”, the disproportionate quantity of reiteration, by comparison with all other prospections in the story, surrounding whether Corley will succeed in bringing ‘it’ off.

That Lenehan’s suspense is ‘released’ we infer from the brilliantly judged final sentence of the story:

A small gold coin shone in the palm.

Here we can again see at work the remarkable phenomenon of inference-based inter-sentential sense-making. For there are none of the explicit links, between this final sentence and earlier ones, that we tend to expect in narrative conclusions. There could have been: much earlier, Lenehan could have been reported asking if Corley was sure he could get the money or get the coin from his woman friend, but the far vaguer are you sure you can bring it off? and work it all right now are all we encounter. The story’s penultimate sentence states that Corley extended a hand so that the word palm in the final sentence is locally metonymically predicted; and the verb shone that appears here was interestingly also used of the lamps in the very opening paragraph of the story. But the specific point that the coin in Corley’s hand is and represents his ‘working it all right’ on Lenehan’s behalf is something that the reader has to infer: it is not textually explicated. So it is the less surprising and perhaps only fitting that the various procedures for identifying high narrativity text have not picked out this final sentence for inclusion in the prospection abridgement.

4. Textualising surprise in “Bliss” and “A Small, Good Thing”

In narrative surprise, it was argued earlier, a script-like important sequence of events, predictable and foreseeable, narratively unmarked, often of a normal kind, emphatically does not continue as expected. Instead it is abruptly displaced by the revelation of something quite unexpected and at least as important, narratively marked and out of the ordinary in the presented situation. Sometimes the revelation is a pleasant or even wonderful one, but just as often it is deeply unwelcome for one or more of the characters (and sometimes for the empathising reader too). As a result, whereas in suspense I suggested that character-focalised evaluations tend
to precede the point of completion where the suspense is released, in situations of surprise character-oriented evaluations tend to follow the surprise-disclosure.

This characterisation may help to explain why it is that surprise seems less reducible to a broad textual template than narrative suspense is. The explanation concerns a paradox at the heart of narrative surprise. On the one hand, it is crucially experienced as a sudden and disturbing shift in the footing or direction of the narrative. On the other hand, as is implied in the idea that a powerful surprise is relevant and potentially foreseeable, there may be, in the prior narrative, textual signals consonant with the surprise that is to emerge. The analyst may and perhaps should find these textual signals. The textual analyst’s difficulties thus relate to the way a writer ‘builds’ a narrative surprise: largely, it is a ‘not building’, a not seeming to prepare, and then the sudden textualising of something very different, and important. So the stylistician needs to focus first on the narration of the surprise, and here corpus methods can help us to demonstrate the differentness, importance, and relevance, relative to prior narrative, characters, and situation, of the passage disclosing the surprise.

Let us begin with Mansfield’s “Bliss”, and with what emerges as in the foreground if the story is ‘abridged’ so as to highlight its putative prospection resources, along the lines prescribed in the previous two chapters. But a preliminary reservation about the abridgement needs to be recorded. This is that it does not yield a significant reduction of the original (it amounts to 70% of the original). This is chiefly because the names of three major characters (Bertha, Harry, and Miss Fulton) are frequent keywords, so that all sentences using their names were retained in the abridgement. This underscores a point made previously, that the eight narrativity parameters and their precise limits, are a provisional model open to adjustment. It should also be noted that the abridgement captures most but not all of Bertha’s FIT reflections.

It is instructive to see how the lexical keyword and cluster analyses that contribute to the abridgement bear on the generation of narrative surprise in the story. The frequent lexical keywords (as noted at the end of 4.1 of Chapter 7) include: pear(-tree) (8, .17%), bliss (6, .13%), and silver (8, .17%). More interesting than considering these prominent lexical keywords in isolation is to consider their immediate textual environment, by examining their concordance lines and looking for potential collocates. This reveals that silver co-occurs frequently with moon, with Miss Fulton’s fillet, and with pear tree and Miss Fulton. For example, sentence 243 (already in the abridgement) runs:

*It [the pear tree] would be silver now, in the light of poor dear Eddie’s moon, silver as Miss Fulton, who sat there turning a tangerine in her slender fingers that were so pale a light seemed to come from them.*
In the case of bliss, its concordance lines show that it is several times preceded by feeling of; the final occurrence of bliss is in this FID sentence (just a few lines before Bertha’s shocked discovery):

Was this what that feeling of bliss had been leading up to?

Concordancing similarly shows that pear tree is almost invariably preceded by lovely and this is tellingly preceded sometimes by your (i.e. Bertha’s) and sometimes by the.

When the story’s multi-word clusters are computed, it emerges that there is only one 4-word cluster that occurs more than twice in the story: into the drawing-room (4 instances). At the 3-word level, by far the most frequent cluster is the drawing-room (9 instances, which of course include the 4 instances of into the drawing-room). And drawing-room, it was noted at the close of 4.1 in Chapter 7, arguably should be listed in its own right as a keyword, depending on the reference corpus one adopts.

The above, then, are some of the characteristics of the ‘automatic’ abridgement of “Bliss”. Returning to the story as a whole and readers’ reactions to it, we can begin by noting its initial schematic scriptedness. What could be more pre-scripted and prospectible than this plot, where a young upper-class woman, blissfully happy with her life and ready now to share this passion and desire with her husband, has her interesting bohemian friends over for dinner but looks forward to the special intimacy of the bedroom at the end of the evening? Everything proceeds to Bertha’s plan, so that even minor variations (such as Harry coming home late) are assimilated. Part of that plan is that at the right time Miss Fulton will give Bertha ‘the sign’ (of special understanding) and this, too, duly occurs, the hostess believes. Particularly script-like is the business of helping guests with their coats and bidding them farewell at the end of the party (there is a comparable scene in “The Dead”). Bertha sees out the Norman Knights, Harry will similarly assist Miss Fulton, and later Bertha will attend to Eddie.

But this bland expectability (so predictable that some at least of it would ordinarily have to be ellipted as unnarratable) is shockingly disrupted when Bertha, herself unobserved, looks from the drawing room into the hall, where Harry is helping Pearl Fulton with her coat. The moment of shock and surprise is carried by a single late paragraph:

While he [Eddie] looked it [a poem] up she [Bertha] turned her head towards the hall. And she saw... Harry with Miss Fulton’s coat in his arms and Miss Fulton with her back turned to him and her head bent. He tossed the coat away, put his hands on her shoulders and turned her violently to him. His lips said: “I adore you,” and Miss Fulton laid her moonbeam fingers on his cheeks and smiled her sleepy smile.
Harry's nostrils quivered; his lips curled back in a hideous grin while he whispered: “Tomorrow,” and with her eyelids Miss Fulton said: “Yes.” (ellipses in original)

How does the abridgement of the story and what it suggests about the story’s narrative progression, together with lexical corpus analysis, help us identify and understand the working of narrative surprise here?

None of the story’s more frequent lexical keywords (pear tree, bliss, silver) or their textual collocates (slender, your, lovely) appear in this crucial but deeply atypical paragraph. There are some words and phrases, besides character names, inherited from the pre-surprise text, such as mention of Miss Fulton’s fingers and smile: earlier she is describes as having a sleepy voice and a strange half-smile that came and went and slender fingers that were so pale. But these have undergone a tonal change: now she has moonbeam fingers and smiled her sleepy smile. We infer a different and ironical kind of meaning to these phrases. One of the textual clues to this is the reporting that Miss Fulton smiled her sleepy smile. Via the redundant local repetition, the formulation implicates some indirect extra meaning: perhaps that Bertha (at last) sees she has misinterpreted Pearl, or that the smile is manipulative or unnatural. Part of the ironising distance is created by the sheer repetition itself, the mention again here of the fingers and the smile, already described earlier: what was previously a descriptive use is in this iteration a more quotative, disengaged mention.

In fact the whole paragraph is in contrast with the rest of the text because it lexically thematises the human body and, especially, the face: arms, back, head, hands, shoulders, lips, fingers, cheeks, smile, nostrils, lips, grin, eyelids. What has Bertha to offer, by way of counterweight? In what terms has she been textually represented? By implication she is a rare fiddle shut up in a case, or a lovely pear tree, silver and slender and vertical and impossibly idealised. Nearly all the body- and face-part items just listed from the surprise paragraph are assigned to the B1 semantic category in the Wmatrix semantic tagset (B1 excludes hands and grin, and includes sleepy). It is clear that there is a ‘spike’ of B1 lexis in just this paragraph of the story: Wmatrix reveals that there are 89 such items in the whole 5,000-word story, and 12 of them feature in this 100-word segment (neither Youmans’ Vocabulary Management Profile software, applied to various text spans, nor Mason’s Paraword captures this specialised kind of lexical density in the story). So this paragraph in the story is all about faces and bodies (but not Bertha’s), other paragraphs hardly at all.

Equally importantly, the ‘surprise’ paragraph is heavy with words and phrases, collocations and colligations, found nowhere in the entire preceding story. Perhaps the crucial sentence, and the most explicit, is:

_His lips said: “I adore you,” and Miss Fulton laid her moonbeam fingers on his cheeks and smiled her sleepy smile._
His lips occurs twice in the paragraph, and only here. Elsewhere in the story we find her lips and, once, their lips, and most importantly and contrastively Bertha’s lips which are seen in the mirror near the story’s outset (s.23):

She hardly dared to look into the cold mirror—*but she did look, and it gave her back a woman, radiant, with smiling, trembling lips, with big, dark eyes and an air of listening, waiting for something...* divine to happen... that she knew must happen... infallibly.

The only adore in the text is Harry’s, in the surprise paragraph, although earlier Bertha has thought how she has an adorable baby. Nowhere else in the text does Miss Fulton lay anything on anything (or give or do anything to anyone), let alone lay her fingers on another person’s cheeks. Not only is she never elsewhere an Actor where another human is Goal, she is rarely in the habit of being Senser of Mental Processes with a human Phenomenon (Halliday and Matthiessen 2004). At one point we are told *Miss Fulton did not look at her [Bertha]; but then she seldom did look at people directly.* She is more in the habit of non-finitely stirring the beautiful red soup or turning a tangerine in her slender fingers. And it hardly needs saying that the Harry who here adores Miss Fulton presents a very different figure from the one whom Bertha earlier recalled had voted her [Pearl Fulton] dullish and cold.

Is there any relevance to the narrative surprise in the fact that just one four-word cluster occurs in the story as many as four times: *into the drawing-room?* Like the *it off all right* in “Two Gallants”, this prominent iteration appears to be far from inconsequential: what might seem like trivial recurrent locative information is crucial to the denouement and the surprise. The last occurrence of the 4-word cluster comes when we are told *When she [Bertha] got back into the drawing-room the others were on the move.* Bertha has just seen the Norman Knights out (“*I’ll come with you into the hall,*” said Bertha). The remaining guests are on the point of leaving, but now when Miss Fulton moves towards the hall to put on her coat we are told *Bertha was following when Harry almost pushed past.* “*Let me help you.*” This is exaggerated attentiveness on Harry’s part, Bertha believes, to compensate for his earlier brusqueness. But the upshot, in terms of character placement in the scene or contextual frame, is that Harry is in the hall with Miss Fulton and the pair believe they cannot be observed by Bertha, who is still in the drawing-room with Eddie. They are mistaken: in search of a poetry book, Bertha has moved noiselessly to a table opposite the drawing-room door. So the prominence by repetition of mentions of the drawing-room is no accident, but instrumental to the plot.

Detailed study of the lexis and syntax of the surprise-recording paragraph in "Bliss", in the original and in the abridgement, shows multiple differences from (and sometimes reversals of) the patterns and emphases previously established in the developing text. In effect the reader is deliberately ‘unprepared’ for this
paragraph by the foregoing text; surprise is therefore almost guaranteed. In Raymond Carver’s story “A Small, Good Thing” there is an example of a narrative surprise which is even more briefly reported than that in “Bliss”, but just as powerful in its context. Equally, drawing particularly on material that is captured in the high-narrativity abridgement, it is built on a textualisation that is in sharp contrast with the prospections created in prior text. The textual patternings at the point of surprise have to be set against the contrasting norms established in the course of the preceding text.

In simplest terms, the preceding abridgement leads the reader to empathise with Howard and Ann, the distraught parents of Scotty, a little boy who on the morning of his eighth birthday is knocked down by a car and after staggering home falls unconscious. For forty-eight hours the parents keep vigil at Scotty’s bedside in the hospital, where the reader at least is in various ways led to expect, despite the suspenseful nature of the boy’s and his parents’ ordeal (marked by many of the features specified in Sections 1 and 3 of this chapter), that eventually Scotty will recover. One instance of cold comfort comes from a brief side-story: Ann has talked to the parents of another boy, who was similarly ill and in need of surgery, and she subsequently learns that the boy has died. Surely, the reader surmises, such tragedy cannot be repeated in the plot? Besides, the doctor assures the parents that Scotty isn’t really in a coma as such. In addition, the story and abridgement are little focused on Scotty, but instead on the parents going through their ordeal of fearful waiting. In the course of this waiting, whenever one of them briefly goes home they are harassed by phone calls “about Scotty” from an unidentified person. The reader knows that this is the baker, whom Ann had visited at the opening of the story, ordering a cake from him for Scotty’s birthday. The baker is angry that no-one has collected the cake or paid him; but Ann is much too preoccupied to remember all this or identify the caller. The baker’s hostile anonymous phone calls create an intermittent second storyline with its own prospections, with Ann and Howard reacting as though cruel harassment is involved, triggering in them a comparably disproportionate response (much later Ann will confess to the baker “I wanted to kill you …I wanted you dead.”). From the reader’s perspective these prospections seem likely to lead only to the most banal of completions: e.g., the parents will remember and pay the baker, or he’ll waive the charges in the circumstances.

There is straightforward suspense then, which the reader carries forward in their processing of following text, as to whether Scotty will recover or not; at the same time many readers report expecting that, in line with recurrent cultural scripts, Scotty will emerge from his ‘sleep’ and be all right. The suspense and the positive expectation are both reflected in what is arguably the story’s key cluster, which happens to be the only 5-word cluster to occur as many as 4 times in the text: going to be all right. All occurrences unsurprisingly occur prior to the passage
of narrative surprise to be quoted shortly below, and they are never modified by a negative; the final instance is typical:

Ann said to the man … “He has a concussion and a little skull fracture, but he’s going to be all right.”

Indeed the simple phrase all right occurs 16 times in the story, nearly always preceded by he’s (denoting Scotty) or be, and none of them is negated. The doctor frequently assures the parents that Scotty is and will be all right, and the parents repeat this to each other. The doctor also reiterates that Scotty is not in a coma – sufficiently often that coma emerges as a lexical keyword (the most frequent lexical keywords however are said, hospital and telephone). Here now is the passage containing the narrative surprise; it occurs when both of Scotty’s parents are at his bedside. The parts of the passage that would enter into the eight-parameter narrativity-foregrounding abridgement are set in bold:

“Look!” Howard said. “Scotty! Look, Ann!” He turned her toward the bed.

The boy had opened his eyes, then closed them. He opened them again now. The eyes stared straight ahead for a minute, then moved slowly in his head until they rested on Howard and Ann, then traveled away again.

“Scotty,” his mother said, moving to the bed.

“Hey, Scott,” his father said. “Hey, son.”

They leaned over the bed. Howard took the child’s hand in his hands and began to pat and squeeze the hand. Ann bent over the boy and kissed his forehead again and again. She put her hands on either side of his face. “Scotty, honey, it’s Mommy and Daddy,” she said. “Scotty?”

The boy looked at them, but without any sign of recognition. Then his mouth opened, his eyes scrunched closed, and he howled until he had no more air in his lungs. His face seemed to relax and soften then. His lips parted as his last breath was puffed through his throat and exhaled gently through the clenched teeth.

The doctors called it a hidden occlusion and said it was a one-in-a-million circumstance.

What does this passage show? Like the paragraph in “Bliss” in which Harry and Pearl embrace, it is lexicogrammatically strikingly at variance with preceding trends and expectations. This is particularly true of the latter half of the passage, from The boy looked at them, but without any sign of recognition onwards, which is arguably the nucleus of the surprise. The surprise is not that the boy and the doctors and the parents should be involved, but that they should be described in ways so different from previously. Nowhere else in the story, for example, is it reported that The boy looked. On the contrary everybody else is reported looking (in both perceptual and equative senses), very frequently: we are told how Ann looked, Howard looked, the doctor looked, and so on. A little earlier, for example, the
following text occurs (where, incidentally, the link between looking and recognition is reflected in Howard’s nod):

She [Ann] went to the door, where she turned and looked back. She looked at the child, and then she looked at the father. Howard nodded.

As a result, looked features as a lexical keyword in the story, with 42 occurrences comprising 0.43% of the text (keyness of 43), in fact the most frequent fully lexical keyword in the story (hospital and telephone, mentioned earlier, are more key but only half as frequent, with 23 and 21 instances respectively).

The without (of without any sign of recognition) by contrast is used here to contribute a message of disengagement that is strikingly congruent with its previous uses in the story. This negation-carrier is used seven times in all, the present instance being the last; here are the relevant concordance lines:

1. older man with a thick neck, listened without saying anything when she told
2. him for his birthday that afternoon. Without looking, the birthday boy step
3. stood at either end of the gurney without saying anything, though once
4. she put on the stand beside the bed. Without a word to them, she took blo
5. to begin. She stood looking at them without saying anything more. She w
6. she passed the cart. She hurried on without looking at any of the nurses a
7. Scotty?” The boy looked at them, but without any sign of recognition. Then

Clearly in the recurrent word-choices to the right of without there is a pronounced association with, and even a prosody of, explicit non-communication: without saying, without looking, without any sign, etc. This seems to be a somewhat specialised association: saying and looking are not prominent collocates of without in the Cobuild Collocates CD. But given the prior co-occurrences of without and non-communication in the story, the reader is arguably primed for Scotty’s non-recognition here; whereas if the text had run The boy looked at them, but without any sign of discomfort, the latter phrase would run entirely counter to the prior pattern.

Limitations of space prevent a detailed charting of how subsequent lexical items relating to the boy and the doctors are similarly used in powerful combination to generate narrative surprise. But the combination blends familiar and foreseeable lexis and grammar (like without, above) with unprepared-for lexis (scrunch, howled, occlusion, one-in-a-million circumstance, all of which are single-occurrence items in the text) and unprepared-for grammar (e.g., the boy as ‘agent’ of mental and verbal processes such as looked and howled, and the colligation of doctors with called it).

Another example of the foreseeable lexis is the word closed. Like looked, this is a lexical keyword in the story, occurring 22 times (keyness of 51). Closed overwhelmingly collocates with his eyes or her eyes, one or other of these occurring
to the left or to the right of the node closed (the Cobuild Collocations CD, records eyes as the top collocate of closed and, interestingly, closed as the top collocate of eyes). As recently as the opening of the ‘surprise’ passage the reader has been told that The boy had opened his eyes, then closed them. So for Scotty’s eyes to be closed is no surprise at all, just part of a familiar coma script. His eyes have been closed since soon after the accident: that has been a key symptom of his illness. But for his eyes to be scrunched closed is unexpected, and may even mislead the reader to interpret this as a conscious act on Scotty’s part. Interpreting the boy’s movements as intentional seems to be immediately supported when the text continues and he howled, howling by children being strongly associated with their fully conscious protesting at pain or discomfort (a child howling after it has suffered an accident will itself be part of a familiar script for many readers). The range of collocates of howled in the Bank of English (where humans howl in pain or in anger, or with laughter) suggests that it usually accompanies heightened, emotionally-charged, consciousness. In the Carver story, however, Scotty’s howling is terminal, and perhaps not strictly howling at all but a reflex expulsion of air with phonation which merely sounds like howling. Thus while howling is foreseeable lexis – we are primed to encounter it in ‘child reacting to and recovering from trauma’ scripts – the reader suddenly grasps that such a script does not fit here. Howled is therefore both expectable and a misrepresentation. Such extremely localised misrepresentations seem to be a feature of narrative surprises: cf. Harry, in “Bliss”, banally undertaking to help Pearl with her coat; or “Indian Camp” where we are told at the beginning of the surprise passage that “The Indian lay with his face toward the wall”.

We have seen that narrativity abridgements tend not to include sentences that release or terminate a suspense (these being a backwards-oriented ‘answering’, i.e., not prospective). Similarly, the abridgements cannot be guaranteed to record the core of a narrative surprise. This is because, by definition, a surprise is largely unprospected, or only quite indirectly prospected. It is created by the relatively abrupt report of events or states that are not evidently repetitions or variations on previous textual material. In Hemingway’s “Indian Camp” the man has been mentioned before, as has his body rolling over in bed, so those elements are repeatable at the point of surprise; but the significant other elements – such as his throat being cut from ear to ear – are not and cannot be repeats. Even though surprises are unprospected, their emergence exploits prospection: the foregrounded seam of most prospected material is used as a springboard for the surprise, since that material creates expectations that turn out either not to be fulfilled, or if fulfilled, done so in ways that are different and perhaps of lesser relevance than the suddenly emergent surprise. And as in the case of the span of text between a suspense-initiating prospection and its delayed completion, the text at and around the point of surprise will be disproportionately dense with mental process projections and modality
(Parameters 5 and 8). It also tends to have a heightened density of narrative actions verbs with main character as participant (Parameters 1 and 2), and disorientating unprospected uses of the story’s lexical keywords and key clusters (Parameter 4).

This chapter has suggested some of the ways that a semi-automatic abridgement of a story, and a variety of corpus analytic methods of exploring the lexical (and to a lesser extent syntactic and semantic) patterning in that abridgement, can bring us closer to a linguistic description and understanding of suspense and surprise, these powerful exploitations of narrative prospection. Suspense and surprise exploit the expected and the unexpected, the former palpably delayed and rendered unlikely, the latter rushed into. But I have argued that the most powerful of these effects in literary stories are not created at the level of paraphrasable events or plot summary. They are deeply rooted in the texture of the story, in an interplay or clash at the level of word-choice and sentence-choice. On the one hand a literary narrative displays established and foregrounded cross-text trends in the realisations of the eight nominated parameters of narrativity. Thus there are text-wide trends in the kinds of lexical prominence, character-action verb combination, negation, character thought, and direct speech prospection. On the other hand there are, across the limited span of a suspenseful section or around the nucleus of a destabilising narrative surprise, multiple local departures from those text-wide norms.

In the present chapter I have begun to put the eight-parameter model of narrative progression – itself a work in progress – to work on some real ‘cases’: stories in which suspense or surprise or both are generated. In the final chapter I discuss how much further such testing can go, the degree to which corpus methods are proving useful, and the need now to do more to connect the postulated textual narrativity ‘triggers’ with readers’ disclosed responses.
CHAPTER 10

Next steps


There is clearly much more to be done in the way of systematic corpus-based analysis and testing, to underwrite claims about which segments of a short story are incontrovertibly in the foreground in the creation of narrativity and narrativity-derived reader reactions. Since this particular kind of study of narrative progression is in its infancy, it seems fitting that this book is so clearly a ‘work in progress’. The story is far from over; there is much more to tell – or for readers and analysts to find out – about what happens in the progress of a story.

What I have chiefly hoped to do is demonstrate the usefulness of some kinds of relatively simple and relatively automatic searches. The searches included ones for kinds of lexical-semantic repetition and recurrent multi-word clustering, and kinds of character-focussed text (identified by keyword-prominent name) or action-focussed text (identified by narrative-tense verbs with main character as participant). There were searches too for kinds of character-internal mental dramatisation (in semi-automatically detectable sentences of free indirect thought), and the expressions of subjectivity and evaluation found in the use of projecting mental process and modal verbs. Further searches attended to kinds of foregrounding by ‘disnarrating’ and negation, and the gapping or secrecy that negation and pronominal indefiniteness can generate (e.g., use of something, a thing, nothing, at points of prominence). Also judged crucial to track were the kinds of event-prospection expressed in characters’ public (direct speech) proposals and requests. Comparatively easy to isolate are the opening sentences of successive narrative paragraphs; these are often crucial in the continual and narratively-significant resetting of place, time and topic.

I have argued that to dwell especially on these parameters of a narrative text is to focus more sharply on a story’s chief resources for creating and developing narrativity or prospection. I have proposed that these features – in their sentential settings – are especially instrumental in holding the reader’s attention on the progression, and in triggering in the reader a range of strong expectations and feelings. Those expectations and feelings are the complex intellectual and emotional reactions, generated in the course of the act of reading and not merely at its end, to
which we give such general descriptive labels as suspense, surprise, mystery, tension, secrecy, happiness, sensuality, injustice, absurdity, cruelty, and fairness.

The kinds of corpus-detectable textual patterning listed above, and not the innumerable other kinds of patterning one might find and comment upon, are postulated as central. And I have sought to specify each parameter in sufficient detail that it can be tested, refined, and if necessary rejected as secondary not primary, and replaced by alternative parameters or criteria – by me and by other researchers. I do not pretend that the parameters are definitive, collectively or individually, or deny that much adjustment and correction of them may be needed – especially if an attempt is made to extend the application of the model to a wider range of literary narratives, from different periods and different genres or subgenres. But I believe we can develop a corpus narrative stylistics that will apply and test such adjustments and corrections. An extreme corollary of this privileging of the nominated eight parameters is the prediction that a modern short story without instantiations of all or most of these parameters would be devoid of narrativity or powers of prospection, and as a consequence would not prompt focussed reader expectations. Modifying texts and designing experiments to test such a claim is not straightforward, but such testing is a clear and potentially rewarding avenue for future research. Such tests are also likely to help in the evaluation of any proposed revisions to the parameters, including narrowing or broadening of their scope.

All eight parameters to which ‘high narrativity prospection’ is attributed were discussed in detail in Chapters 7 and 8, for which the first six chapters were a preparation; and some testing of the model on selected stories was undertaken in Chapter 9. But this too needs to be greatly enlarged.

Throughout, a strong preference was for a relatively shallow text searching, hence one that is more conducive to automatisation or semi-automatised implementation, rather than nominating and attempting to search for instances of rather more abstract or covert features. Once a corpus-analytic approach is adopted, there is a self-fulfilling side to this preference for textually-explicit features (such as the initial sentence in each paragraph of narration) rather than ones that, though semantically rich, tend to be more implicit and open to identification disputes (e.g., textual expressions of heightened character emotion): computational methods are enormously more reliable in identifying the former rather than the latter. Besides, even if it were possible reliably to identify all and only passages and phrases of heightened character emotion, this would prompt methodological objections. It would be suspected that the analyst had selected “for special attention” textual material that was already by its identified nature of a special status. Subsequent analytical ‘findings’ would be resting tacitly on a powerful semantic-evaluative pre-selection. By contrast there is no such preliminary special meaning that accrues to the first sentence of a paragraph simply by virtue of its being the first
sentence. The selection criterion is purely formal or positional, not semantic. First sentences of narrative paragraphs in literary stories could in principle frequently be of the order of Then a very strange thing happened or Now I will explain what all this was leading up to, which would provide slim pickings for narrative prospection. Like the use of negative free morphemes and proposal-projecting direct speech and all the other specified parameters, the first sentence in a narrative paragraph is a narrative discourse resource that can be deployed in a multiplicity of ways to carry an infinite variety of semantic or evaluative indications. But we have found that in practice, for the kinds of stories analysed, this resource tends to be centrally contributory to narrative prospection. The beauty of such literary linguistic claims is that you do not have to (you must not!) believe the analyst: you can easily test these things for yourself. For example you could make an abridgement taking one sentence from the middle of each narrative paragraph of a story; or you could make a sequence of all the final sentences in those paragraphs. You could then evaluate these abridgements, or better yet have readers judge them, for narrativity-informativeness by comparison with the abridgement based on paragraph-initial sentences.

The textually explicit or routinely identifiable is inevitably more congenial to the stylistician, who believes in the potential significance of every aspect of the crafted ‘surface’ of the text, than is material that is covert or contentiously derived. And the more obvious and identifiable these elements are to a machine following fairly simple instructions, the more reasonable it is to suppose that they might equally be noticed by the reader – indeed, that for the reader they might be unignorable. The claim needs to be made with care: it is not that textual obviousness alone is sufficient condition for something to be likely to be psychologically real for readers. It is clearly possible, for example, for a simple mechanical computation to identify and retrieve a ‘surface’ text feature (e.g. all words, or all sentences, containing the letter q) without it being at all obvious that the feature is psychologically real for readers. So textual obviousness is a necessary condition but not a sufficient one; it is invoked together with varied others. In the present study, a further independent requirement has been that features be ‘harvested’ and assayed within the whole sentence (at least the whole matrix clause) in which they occur.

Those, then, have been the key assumptions motivating the present study. It is for readers finally to decide whether the textual distinctions they have given rise to – in particular the isolation of a narrativity abridgement – do indeed bring us closer to an understanding of the grammar of short story narrativity and prospection.

As regards story-bridgements themselves, it must be emphasised that these were only created and examined independently of their fuller surrounding text so as to try to test the claim that their content was indeed crucial and shaping of prospection. This required some form of separation of the narrativity sentences
from the potentially thesis-confounding remainder. But again from the point of view of psychological reality and the reader’s processing of the text, these abridgements have no separate existence; it is better to think of them as an intermittent chain of narrativity-crucial sentences distributed along the length of the story. A loose analogy might be to a long-distance walking route, which intermittently takes the hiker above the treeline. The ‘above the treeline’ stretches of the walk may be the most exhilarating, affording the best views of distant prospects, but there is no possibility of walking the route in such a way that one experiences only the above the treeline sections: they have to be encountered in due course, interspersed experientially with the walking of the lower stretches of the path.

2. Expanding and refining the model: modality and evaluation

In a book of this length, devoted to contributing to a new mode of narrative analysis, many important issues and complications have had to be silently passed over in order to focus on what were regarded as core questions. But if this form of analysis is to develop and become established, proper note will have to be taken of the topics I will briefly mention below.

A vast area of narrative complexity that has been given only limited treatment in this study concerns modality and evaluation, and the reliability or uncertainty with which whole episodes or particular facts are narrated. Even while accepting the general fictionality of a folk tale or a short story, the reader needs to make subtle calibrations as to the plausibility of the narrated situation, the reliability of external and internal narrators, and the satisfactoriness of any fanciful departures from the mundane. And beyond such broad assessments of reliability, a sophisticated literary text often affords the reader continual indications of the desirability, certainty, or otherwise (according to a focalising character, an intradiegetic or an extradi- egetic narrator), of the events, actions and states which are reported. Any text may contain a variety of textual signalings which qualify, evaluate and appraise (Hunston and Thompson 2000; Martin and White 2005) what is reported. Their com- parative neglect in the present study is I believe justified, since the focus has been on narrativity and expectation, rather than every aspect of narrative texture, and the evaluation embedded in the telling plays a secondary role in the creation of expected progression. Still, some consideration of modality and evaluation is essential, and this was given in such parameters as those attending to character-inte- rnal thought, modal and mental processing verbs, and direct speech prospection.

A close partner of evaluation, in text construction and construal, is perspec- tive or point of view. In narratology, the point of view from which events and the characters populating them are described is as important as the events themselves,
generating detailed studies of a phenomenon sometimes called focalisation. The present study has paid little direct attention to point of view or perspective. Everything to do with perspective makes assessments about narrative prospection more complicated, more conditional upon how reliable we consider the perspective and source to be, from which matters are told. As Jahn (1999: 95) puts it, focalisation can be understood as “providing and managing windows into the narrative world, and of regulating (guiding, manipulating) readerly imaginary perception”.

For a corpus stylistic approach to narrative to be more complete and balanced, it will have to incorporate more attention to the perspective from which events are reported, since we know that some perspective (even if only that of an ‘omniscient’ narrator) is always required, and every perspective is partial. Consider, for example, a narrativity abridgement that, if attributable to a reasonable and reliable narrator, would trigger powerful reader reactions of suspense and anxiety. If the narrating perspective is that of an unreliable fantasising character however, then those textual effects and reader reactions are unlikely to take hold. In the present study, stories have been explored where such thoroughgoing faultiness of perspective is not apparent, even though the character from whose point of view much of each story is told (e.g. Lenehan; Bertha; and the parents in “A Small, Good Thing”) tends to be a fallible one, with limitations to their insight on events as they unfold: those limitations are not so thoroughgoing as to prompt readers, in mid-story, to abandon the forming of normal expectations about the story progression and outcome.

In trying to analyse these short stories as progressively and incrementally encountered texts in the way that they are experienced by a reader, I have also adopted quite different methods from those of Phelan’s ethics-oriented and affective rhetorical narratology (Phelan 2005). Phelan conceives of narrative not as a structure but as a rhetorical act; but his interest in the rhetoric of narrative and narration means that, like me, he is especially interested in narrative progression. His work should prompt us to begin to pay more attention to other kinds of emotional response to stories than such more directly apparent effects as suspense, surprise, tension, mystery, and dissatisfaction. How does a story generate pity, or grief, or shame, in a reader? Phelan writes:

Rhetorical ethics views the ethical dimension of literary experience as part and parcel of aesthetics…. the method of rhetorical ethics is often to approach the ethical dimension through an analysis of those aesthetics – including both a text’s design and the consequences of that design for the affective response of its audience. (Phelan 2004: 647–8)

Attempting to link such an emphasis on the ethical dimension to corpus-informed stylistic evidence of the kind focussed upon in the present approach is a major challenge. Ultimately, since both approaches are at base devoted to a better
understanding of readers’ experience of the narrative, with what the text does to them and how they respond to the text as their objects of enquiry, there simply must be a link; or many links.

In future studies it will be important too to continue to attend to the still-developing models of discourse and language of corpus linguists like Hoey, Louw, and the late John Sinclair. Hoey’s work has been particularly influential on the present study, and his new theory of priming promises to be equally important (Hoey 2005). This theory broadens out the foundational notion of collocation to argue that, at least within particular genres, some words tend to collocate, collage, and associate semantically with particular kinds of partner material, and that they in addition tend to occur in particular sentential and textual positions – i.e., that there are weighted tendencies at almost every level of textlinguistic classification. This exciting contribution to collocational and language theory merits further scrutiny to see how far it is valid and applicable to, for example, the recurrent vocabulary of narratives. In Hoey’s priming model it is stressed that the priming is in the language user, not strictly in the language, and therefore varies from user to user, reflecting their prior exposure to kinds of language behaviour. Some have questioned Hoey’s adoption of the term priming, which has a long history of use in psycholinguistic experimentation, since in practice all the evidence pertains, as in most collocational corpus linguistics, to disproportionate frequency or recurrent pattern. An alternative less technical-sounding term for the phenomena that priming theory attends to would be expectation.

Just how applicable priming theory will be to the patternings we find in literary texts remains to be seen. Some commentators have been doubtful (e.g. Dillon 2006). And as already noted in this chapter, narratives present particular complications because they are so chameleon-like in their shifting adoption of point of view (by comparison with, e.g., a typical newspaper news article). This is what leads Wikberg to comment (2007: 93):

A question that arises in fiction is to what extent lexical priming is to be looked at from the point of view of the author or possibly the characters. When authors individualize their characters, this has to be realized in terms of what is done, said, thought, intended, or hypothesized.

Another reference point for this section, which has reflected on some possible future directions of corpus stylistic research on narrativity, returns us to one of the deeper questions that arises with the use of corpus methods of text analysis at all. This concerns the nature of ‘text’ and the effects and purposes of visualisation and computer-aided text analysis. Michael Barlow (2006) has commented interestingly on these issues. His remarks come at the end of a paper in which he has displayed various patterns in graphical forms, the re-presenting of textual evidence that
language analysts routinely perform whenever they put text fragments into Powerpoint slides or into a textual figure in a chapter or article. These visualisation means should never be mistaken for the end of the analysis, Barlow reminds us:

In presenting the searches and graphical representations above, I may have given the impression that what I am proposing is a technology project, dealing essentially with technical or computational problems. This is not at all the case. I do not need time and space and a community to work out how to get a text to show hapax legomena in a red font. That is not something I have to grapple with. The problems to be faced are conceptual, not computational. What, for instance, is a text? How should a text be represented visually: as a two-dimensional physical object, perhaps, with representations of pages; or as a uni-dimensional object consisting of a string of characters; or as an abstract object viewed in terms of relations among words or relations among words and annotations? I have to evaluate the consequences of adopting different models and while I do not necessarily have to adopt a single representation of a text, I do have to assess the usefulness of visualisations based on different ‘views’ of a text.

These points always need to be kept in mind, not least when foregrounding and displaying kinds of high-prospectivity sentences in short stories.

3. The reader’s experience of the text

In this study of progression, I have tried to indicate at least some of the links with the larger assumed picture of what is undergone when a person reads a literary narrative. Among the factors most relevant here, repeatedly recognised in the reading of literary narrative, are immersion, expectations, thoughts and emotions. Part of the power of literary narratives, despite their fictional characters and situations and absence of direct real-world consequences, stems from their ability to absorb a reader’s full attention, to the point that real-time obligations and concerns are temporarily forgotten. The metaphor of immersion is often invoked. It is because the reader may be so thoroughly ‘taken into’ the textualised narrative situation, so caught up in the web of wordings, that the kind of attention to textual signallings here argued for has some plausibility and point.

The immersion in story is, to continue the metaphor, not a question of plunging into a static pond or bath, and this has enormous consequences for the matter of expectation. After the first few paragraphs of a story the reader is in one state of immersion, but by the later stages of the story, many pages of shifting events and states later, the conditions of immersion will inevitably be different. So it is rather as if the reader enters a fast-flowing river or sea at one place and is carried by powerful forces (the textual material, encountered sequentially rather than all at once)
to another point. With each perceived shift in the conditions of immersion, such as perceived shifts in the place of immersion, prior expectations may be strengthened, or met, or set aside, and new expectations can emerge.

While ‘expectations’ is a general term for narrativity-oriented reader responses, slightly more particularising ones often invoked include thoughts and emotions. And there is a growing interest in both thoughts and emotions as key parts of the array of responses or effects that readers attest experiencing during and at the close of literary narratives (Hogan 2003, Miall 2006, Oatley and Johnson-Laird 1987, Oatley 2003). At the same time identifying and delimiting particular thoughts and emotions, as distinct from readers’ reports of having thoughts and of feeling emotions, is obviously fraught with difficulties. As far as the present study’s treatment of narrative suspense and surprise is concerned, it is evident that both the thoughts and the emotions that may accompany those reactions are complex phenomena, a combination of ideational and affective responses to text, where a secure separation of what is thought from what is felt may be impossible. Still, the terms in which readers discuss and comment upon their own incremental processing of literary stories is clearly an area which may come to enrich the present study’s focus on the textual resources (see, e.g., Fialho 2007).

4. Directions for future research

Future work testing the narrative prospection model against readers’ responses to story sequence will have to include attempts at matching corpus linguistic identifications of the eight narrativity-foregrounding features (in story text segments, for example, of just a few paragraphs in extent) against readers’ elicited brief responses regarding plot and character expectations. As more segments of the story are analysed (by the text-analytic software or by the reader), it can be postulated that a cumulative network of actual and potential phraseology emerges and strengthens; in theory, expectations should become more secure and reliable.

Just how, in a principled way, such a developing network can be ‘pruned’ of less relevant lexis, and how a hierarchy of relevance might operate (‘promoting’ certain lexical items and chains as prospection-crucial, carrying the main events and states, while demoting others), are questions to be grappled with in future studies. Essentially, can the prospection abridgements be made shorter and even more focussed on progression? To some extent, but only to some extent, the questions are less pressing in the case of short stories, where the array of keywords, clusters and lexical chains are inevitably less complex than in longer narratives. But it is assumed that the shaping of hierarchies of narrativity relevance must be an important activity in the reading of novels and similar extended narratives, and
Chapter 10. Next steps

this is presumably governed by considerations other than phraseological or synonymous repetition. Neither frequency nor repetition alone is a sufficient or decisive consideration.

Very much to the contrary, some analysts have argued that the single mention of a particular word in a literary narrative may be felt to be more powerful than numerous multiply repeated items; if this is so, we must look to stylistics, narratology and corpus methods the better to help us to understand how this is possible. Since the majority of words in a text tend to be single instantiations, specific grounds need to be found for distinguishing the few resonant or thematic single lexical instances from the bulk of *hapax legomena*. And beyond the putatively significant single instance lies the potential of un-named entities to wield powers of prospection and narrativity: words evoked or implied but never actually inscribed in the text. While not denying this as a possibility, the position maintained here is that such textually absent words (which might well contribute to a thematic analysis of the text) need to be warranted by the wordings actually present in the text, for their postulation to be justified in a stylistic or textlinguistic approach to narrativity. With words important to a text but not occurring in that text one is at the border of stylistic analysis.

Future research in this area will also have to attend in detail to the question of the extent that external real-world and story-schema knowledge aids progression expectations. That such knowledge informs the reading of narratives, and therefore very likely informs progression expectations, is undeniable; and it has been extensively explored in psychological studies of narrative comprehension. But how it does, and how its role may vary from reader to reader, is immensely difficult to measure. It is, for instance, impossible to present the same story to a reader reading in the two relevant conditions: with access to real-world and schema knowledge, and without any such access. A beginning will have to be made at answering these questions in relation to the textual prospection model proposed here, initially through interviews and questionnaires and similar probes of reader knowledge, but directed to those kinds of background knowledge and assumptions that appear likely to be most crucial to the story under scrutiny.

How can readers best help us advance our understanding of the linguistic bases of narrative prospection? One experimental strategy would be to impose some degree of ‘phasing’ on the reader’s processing of the story, by presenting the target story in pre-determined sections or segments – perhaps of the kind of length and with the story-sensitive boundaries that were selected in Chapter 6, where to obtain a more delicate keyword analysis “The Love of a Good Woman” was divided into 26 sections. Alternatively a story can be more mechanically subdivided into segments of a relatively constant length (perhaps then shortened or lengthened so as to end at the nearest paragraph break) of, for example, approximately 200 words.
Adopting a fairly well-established think-aloud protocol, readers’ immediate oral responses (during reading and for a limited time – no more than a few minutes immediately afterwards) can be solicited and recorded, concerning what they judge the segment to be about and what it makes them think the story’s continuation and outcome will be. One adjustment of this basic method to be explored is whether different and perhaps more interesting results emerge if readers are tested in pairs, one participant being asked to read and then explain a story to their partner, the story being presented and relayed by the participant phase by phase. With roles reversed, readers’ progressive story uptake and expectations can be probed using a second story (similarly unknown to both participants). There are many other possible research strategies (see, e.g., the discussion in Chapter 4 of van Peer, Zyngier, and Hakemulder 2007).

Future work will also provide the kinds of findings about the most convincing textual profiles for prospection, which may lead in turn to revisions to the eight narrativity parameters I have proposed. Many of those parameters are open to kinds of major or minor revision; one or two – e.g., the first sentence of narrative paragraphs – have to be accepted or rejected in a more absolute way. But those amenable to revision or adjustment, involving a narrowing or broadening of their scope, most obviously include the following:

a. Should the parameter of foregrounding by virtue of character-naming keyword sentences be confined to the topmost keyword-named character only, or extend to other frequent keyword-named characters?

b. This suggestion would move the model further in the direction of selective attention to ‘negation-carrying material’ than the suggestion, in Chapter 9, that only lexicalised negation be attended to. Within the lexicalised instances, might attention be better limited to sentences containing free-morpheme negation narrowly defined (e.g., no, not, never, nothing) rather than more broadly so (e.g., also including sentences containing without, fail to, nearly, lacking, etc)?

c. Should the FIT probe include after all, under Rule 1, sentences where adjacent to the modal and pronoun is n’t or not (rather than the differential treatment of these negative particles alone, in uninverted pronoun + modal instances, as proposed in Chapter 8, Section 1.5)?

d. How wide should the range of attention to modals and mental process verbs extend?

At the same time, the testing of corpus linguistic resources must continue: besides concordancing, collocation and keywords, the present study found clusters potentially more significant to literary interests than might have been expected. Interestingly, the latest version of Scott’s WordSmith Tools enables the researcher to
explore a text’s ‘concgrams’. A ‘concgram’ is a listing of all instances of proximate co-occurrence, in any order, of two or more words that are of interest. Besides Hoey’s priming proposals, it may be that some systematic adaptation of the semantic prosody notion (Sinclair 2004, Louw 1993, Hunston 2007, Whitsitt 2005) will prove valuable in textual analysis of narrativity.

Another kind of revision to the model needs to explore further whether the narrativity features are best treated equally, on all occasions of their occurrence, or whether there should be greater editing of the notional abridgement, after the initial retrieval, so as properly to attend to ‘hotspots’ of narrativity. In the multi-parameter narrativity derivations, it was noted at several points in the discussion that the occurrences of a particular feature were not evenly distributed across a text, but tended to cluster and concentrate, often doing so in a few sentences where a multiplicity of such features all seemed to be particularly prominent and where a narrativity peak or heightened prospection intuitively seems to be involved. Working on rather different materials, using somewhat different parameters, and with different methodological goals, these are topics on which many others have commented (see for example the pioneering contributions of Longacre 1983 and Labov 1972). The implications of these narrativity feature clusterings in literary stories, perhaps marking kinds of major phase-completion in stories, are worth much fuller exploration.

Since this study is a work in progress, and has here reached the end of a beginning rather than any more advanced point, it is appropriate that it has no resonant conclusion. Have we come far in the search for textual prospection guidance, the textual roots of suspense and surprise? It is up to the readers of this book to decide. But in doing so they must consider whether there is anything intrinsically defective in the methods adopted. My own view is that there is not, and that to the extent that narrative texts defy ‘capture’ of all their strands of narrativity, and to the extent that subsequent text can flout established expectations and prediction, these are reflections of the system-breaching indeterminacy of language itself, in form as in meaning. It is in the nature of discourse, language-in-context-of-use, both to enable partial prediction of the discoursally subsequent and to defy total predictability. This is a freedom in language, and in the purposes to which language is put (including narrative purposes), that resists reduction to a systematised analysis.

But it does not follow that all analysis is therefore doomed or worthless – half a loaf is better than no bread. Nor, for that matter, is analysis something that readers and linguists can avoid doing: such reflexivity in our dealings with our language is a central property of human language, according to integrational linguistic thinking (Harris 1998). What is the alternative to an analysis of text, of the general kind proposed here, in which some aspects are claimed to be more important than others for particular effects (prospection/narrativity)? It is to assume
that everything in a text is equally important for all purposes (for narrativity, characterisation, pace, tone, etc.); it is therefore to reject the validity of analysis or intratextual discriminations altogether, and to argue that readers and text researchers can only safely make holistic comments about overall readings and reactions. But that would fly in the face of readers’ intuitions and discursive practices. Readers frequently comment on the felt effects of particular 'bits' of narratives, to say nothing of the observations and practices of writers. The present methodology may be criticised as to the particular selection of narrativity criteria; but to reject the very idea of identifying processes of selection, and of postulating the presence of textual foregrounding or instrumentality with regard to prospection would be to deny the possibility of most kinds of narrative analysis altogether.

Sonia Zyngier, one of my wonderfully helpful editors, has suggested it might be fitting to end with a quotation from a short story—a quotation that might speak to the book's concerns. Well, there are any number of final paragraphs, in different stories from the incomparable Alice Munro, that, to a degree, might suit—and why it is that the endings of the stories seem especially suitable for quoting is an interesting question in itself. But none of these quotations would suit perfectly, because each is so completely integrated and made significant relative to the entire story in which it occurs. That is the paradox in attempting a stylistic analysis of one dimension (narrative progression and reader expectation) of a perfectly integrated text. The existence of stories with this perfect integration of effects is the motivation (no lesser source of motivation would suffice!) for the attempted analytical dissection, which their integrated unity defies. In a suitably celebratory review of Munro's art, Jonathan Franzen similarly finds, when he tries to extract a 'capsule' quotation from a Munro story, that he can't stop there (on ethical grounds, it seems): “I want to keep quoting, and not just little bits but whole passages”, even the whole text. So rather than a Munro quotation, it might be better to end with a Franzen (2004) quotation on Munro. He writes:

The [narrative] moments she's pursuing now [in her later work] ... are moments of fateful, irreparable, dramatic action. And what this means for the reader is you can't even begin to guess at a story's meaning until you've followed every twist; it's always the last page or two that switches all lights on. ...For as long as I'm immersed in a Munro story, I am according to an entirely make-believe character the kind of solemn respect and quiet rooting interest that I accord myself in my better moments as a human being.

That is the quality of experience that the finest literary stories engender, and this is what justifies the attempt to pinpoint those stories’ textual characteristics, by all empirical means, including corpus linguistic ones.
References

Cobuild Concordance and Collocations Sampler. (http://www.collins.co.uk/Corpus/Corpus-Search.aspx last consulted 31/1/07).


Scott, M. & Tribble, C. 2006. Textual Patterns: Key Words and Corpus Analysis in Language Education. Amsterdam: John Benjamins.


<table>
<thead>
<tr>
<th>Name</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andersen</td>
<td>150</td>
</tr>
<tr>
<td>Aristotle</td>
<td>97</td>
</tr>
<tr>
<td>Barth</td>
<td>150</td>
</tr>
<tr>
<td>Barthes</td>
<td>167</td>
</tr>
<tr>
<td>Barlow</td>
<td>41</td>
</tr>
<tr>
<td>Bowen</td>
<td>44</td>
</tr>
<tr>
<td>Bramlett</td>
<td>144</td>
</tr>
<tr>
<td>Butler</td>
<td>C. 19</td>
</tr>
<tr>
<td>Carver</td>
<td>26–27, 63, 144, 169, 184–187</td>
</tr>
<tr>
<td>Chafe</td>
<td>9, 126</td>
</tr>
<tr>
<td>Cook</td>
<td>94</td>
</tr>
<tr>
<td>Coulthard</td>
<td>15</td>
</tr>
<tr>
<td>D</td>
<td>194</td>
</tr>
<tr>
<td>Douthwaite</td>
<td>71</td>
</tr>
<tr>
<td>Eggins</td>
<td>146</td>
</tr>
<tr>
<td>Emmott</td>
<td>37, 64, 126</td>
</tr>
<tr>
<td>Fehr</td>
<td>144</td>
</tr>
<tr>
<td>Fialho</td>
<td>196</td>
</tr>
<tr>
<td>Firth</td>
<td>16, 19–20</td>
</tr>
<tr>
<td>Franzen</td>
<td>200</td>
</tr>
<tr>
<td>Grice</td>
<td>7, 46–47, 89, 115, 137, 177</td>
</tr>
<tr>
<td>Hakemulder</td>
<td>71, 198</td>
</tr>
<tr>
<td>Halliday</td>
<td>15, 77, 114, 145, 152, 183</td>
</tr>
<tr>
<td>Hasan</td>
<td>15</td>
</tr>
<tr>
<td>Harris</td>
<td>9, 199</td>
</tr>
<tr>
<td>Hemingway</td>
<td>144, 171, 187</td>
</tr>
<tr>
<td>Herman</td>
<td>95</td>
</tr>
<tr>
<td>Hidalgo Downing</td>
<td>148</td>
</tr>
<tr>
<td>Hoekken</td>
<td>171</td>
</tr>
<tr>
<td>Hoey</td>
<td>16–19, 29, 35, 45, 86, 100, 116, 126, 194</td>
</tr>
<tr>
<td>Hogan</td>
<td>196</td>
</tr>
<tr>
<td>Hunston</td>
<td>192, 199</td>
</tr>
<tr>
<td>Jahn</td>
<td>144, 193</td>
</tr>
<tr>
<td>Jakobson</td>
<td>97</td>
</tr>
<tr>
<td>Johnson-Laird</td>
<td>196</td>
</tr>
<tr>
<td>Joyce</td>
<td>32, 144, 165</td>
</tr>
<tr>
<td>Kipling</td>
<td>144, 169</td>
</tr>
<tr>
<td>Kjellmer</td>
<td>92</td>
</tr>
<tr>
<td>Kuiken</td>
<td>4</td>
</tr>
<tr>
<td>Labov</td>
<td>82, 83, 114, 168, 199</td>
</tr>
<tr>
<td>Larkin</td>
<td>11</td>
</tr>
<tr>
<td>Leech</td>
<td>150</td>
</tr>
<tr>
<td>Lodge</td>
<td>172</td>
</tr>
<tr>
<td>Longacre</td>
<td>114, 199</td>
</tr>
<tr>
<td>Louw</td>
<td>20–22, 33, 44, 46, 194, 199</td>
</tr>
<tr>
<td>Mahlberg</td>
<td>22</td>
</tr>
<tr>
<td>Mansfield</td>
<td>116, 144, 170</td>
</tr>
<tr>
<td>Martin</td>
<td>192</td>
</tr>
<tr>
<td>Mason</td>
<td>37, 182</td>
</tr>
<tr>
<td>Matthiessen</td>
<td>77, 152, 183</td>
</tr>
<tr>
<td>McKoon</td>
<td>170</td>
</tr>
<tr>
<td>Miall</td>
<td>4, 196</td>
</tr>
<tr>
<td>Moon</td>
<td>25, 131</td>
</tr>
<tr>
<td>Morrow</td>
<td>126</td>
</tr>
<tr>
<td>Munro</td>
<td>77, 200</td>
</tr>
<tr>
<td>Mukařovsky</td>
<td>71</td>
</tr>
<tr>
<td>Nørgaard</td>
<td>34, 148</td>
</tr>
<tr>
<td>Oatley</td>
<td>196</td>
</tr>
<tr>
<td>Palmer</td>
<td>144</td>
</tr>
<tr>
<td>van Peer</td>
<td>71, 171, 198</td>
</tr>
<tr>
<td>Phelan</td>
<td>193</td>
</tr>
<tr>
<td>Prince</td>
<td>148, 175</td>
</tr>
<tr>
<td>Proust</td>
<td>173</td>
</tr>
<tr>
<td>Raabe</td>
<td>144</td>
</tr>
<tr>
<td>Ratcliff</td>
<td>170</td>
</tr>
<tr>
<td>Rayson</td>
<td>4, 49, 107</td>
</tr>
<tr>
<td>Sanford</td>
<td>126</td>
</tr>
<tr>
<td>Scott</td>
<td>4, 41, 48, 198</td>
</tr>
<tr>
<td>Semino</td>
<td>137</td>
</tr>
<tr>
<td>Short</td>
<td>137, 151</td>
</tr>
<tr>
<td>Sinclair</td>
<td>15–16, 42, 44, 114, 129, 177, 194, 199</td>
</tr>
<tr>
<td>Slade</td>
<td>146</td>
</tr>
<tr>
<td>Sperber</td>
<td>46</td>
</tr>
<tr>
<td>Stubbs</td>
<td>20, 24, 95, 150</td>
</tr>
<tr>
<td>Teubert</td>
<td>21</td>
</tr>
<tr>
<td>Thompson</td>
<td>192</td>
</tr>
<tr>
<td>Todorov</td>
<td>5</td>
</tr>
<tr>
<td>Tognini-Bonelli</td>
<td>25</td>
</tr>
<tr>
<td>Töibin</td>
<td>169</td>
</tr>
<tr>
<td>Toolan</td>
<td>30, 34</td>
</tr>
<tr>
<td>Tribble</td>
<td>48</td>
</tr>
<tr>
<td>Updike</td>
<td>59</td>
</tr>
<tr>
<td>van Vliet</td>
<td>171</td>
</tr>
<tr>
<td>Wallace</td>
<td>43</td>
</tr>
<tr>
<td>Werth</td>
<td>37, 121, 148</td>
</tr>
<tr>
<td>White</td>
<td>192</td>
</tr>
<tr>
<td>Whitsitt</td>
<td>199</td>
</tr>
<tr>
<td>Wikberg</td>
<td>194</td>
</tr>
<tr>
<td>Wilson</td>
<td>46</td>
</tr>
<tr>
<td>Woolf</td>
<td>144</td>
</tr>
<tr>
<td>Yevseyev</td>
<td>95</td>
</tr>
<tr>
<td>Youmans</td>
<td>4, 38–40, 182</td>
</tr>
<tr>
<td>Zyngier</td>
<td>71, 198, 200</td>
</tr>
</tbody>
</table>
Topic index

A
Abdication and summary contrasted 72–75
Abridgement, production of 157–163
Action verbs, density of in narration 123–124
“A & P” 59–63
Affective narratology 193
Analysis, necessity of 199–200
“A Small, Good Thing” 169, 184–187

B
Baggage 7
Bank of English, Cobuild’s 45, 90, 131, 187
“The Bath” 169
“Bliss” 70, 116–119, 120–121, 131–132, 141–143, 170, 180
“Boxes” 63–66, 72, 74
British National Corpus 45, 131–132

C
“Cathedral” 66–68, 74–75, 119
Clusters 40–42, 132–133, 183
Cobuild Collocations CD 90–91, 94, 128, 188
Cobuild Concordance and Collocations Sampler 92
Coherence in top keyword abridgements 56–59, 61–63, 68
Cohesion 15
Cohesive chains 116–119
Coh-metrix 29–30
Collocate 31
Collocation 18–21
Concgrams 199
Corpus stylistics 3–4

D
“The Dead” 115, 169, 172
Death, narrative salience of 84
Discoursal suspense 165
Disnarrated, the 148, 189

E
Echoism, long-distance 103–105
“Eveline” 70, 115, 119
Expected and the unexpected, the 8–9

F
Foregrounding 71, 190–192
Foreseeability 167–168
Free Indirect Discourse 102, 144
Free Indirect Thought 135–145, 158
Modal verbs in 138–139
‘frontloading’ vs endloading of top keyword sentences 74–76

G
Generic sentences in narration 82, 84
Genre 5–6, 23
Grammatical suspense 165

H
Heart of Darkness 95

I
Idiolect and keywords 100–1
Idiom principle, the 16
Immersion, readerly 30
“Indian Camp” 171–172, 187
Inferencing in narrative comprehension 170–171
Inquit verbs and narrativity 120–121

K
Keyword verbs 95
Keywords, collocates of 90–91
Keywords, occurrence of 50–51, 93
Keywords, negative grammatical 101–102
Keywords, top lexical 87–90, 128–133, 186–187

L
Latent Semantic Analysis 30
Lexical co-occurrence, rarity in 43–5
Lexical innovation, cumulative 37
Literary uniqueness 23, 28, 43–7
“Little Things” 169
“The Love of a Good Woman” 70, 78–95, 149

M
“Mary Postgate” 169
Mental picturing 84–5
Mental process projection 151–152
Narrativity hotspots 199
parameters, provisionality of 161–162, 198–199
Negation, morphological vs. lexical 149

O
“The Old Man and the Sea” 115
Open choice principle, the 16
Openings and remainders, narrative 108–112
Overt vs. embedded textual signalling 135–136

P
Para-repetition 103–112
Paraword 37–8, 182
‘Phrase’ vs ‘chunk’ 23–24
Plotlinks 97, 99
Point of view, in narration 192–194
Priming, lexicogrammatical 16–18, 194
“The Princess and the Pea” 150

Keywords, WordSmith’s 31, 47–51
Pronouns, indefinite 189
Pronominalized character-reference and narrative saliency 116–119
Proper name keywords 53–59, 70–71
Prospection vs expectation 113

R
Reader response 1–2, 4, 61–63, 80–82, 195–198
Reference corpus, factors in choice of 25–7
“Royal Beatings” 70

S
Schemas and scripts 94, 181
Semantic prosody 20–21
Semantic tagging, automated 107
Sentence, graphological as unit of analysis 113–114

Specificity in narratively-prospective entities/characters 83, 109–110
Speech moves 145–147
Suspense and surprise reversals 173

T
Tellability 86
Temporal markers in suspenseful text 177–178
Text phases as implicit question-answer sequences 86
Textual implicitness vs explicitness 5, 191
Textual prediction 11–12
type-token ratio 34–35

The background 5–7
Top keyword sentences as waymarking 72–6
Transitivity analysis 183

U
Ulysses 165

V
Visualisation 195
Vocabulary Management Profile 38–40, 182

W
Wmatrix 49, 107, 120, 143, 151, 160–161, 182
Wordlist, WordSmith’s 33, 48, 67, 131
WordSmith Tools 143, 198
World knowledge 6
In the series *Linguistic Approaches to Literature* the following titles have been published thus far or are scheduled for publication:

3  **McINTYRE, Dan**: Point of View in Plays. A cognitive stylistic approach to viewpoint in drama and other text-types. 2006. xii, 203 pp.