Michael Toolan

Narrative Progression in the Short Story: First Steps in a Corpus Stylistic Approach

FRAMING ASSUMPTIONS:
PROSPECTION, EXPECTATION, RESPONSE

I am interested in the putative textual signalings of narrative progression, and thereafter the reader expectations that these foster; I am trying to identify such signalings (or narrative prospection, as it is also called) with new research methods, namely those of corpus linguistics. Research of this kind, blending a literary interest with use of corpus tools, is coming to be known as corpus stylistics (or more narrowly, corpus narratology). For readers of this journal I assume that neither explaining nor justifying an interest in narrative progression is necessary, so I will discuss this relatively briefly. I will spend a little more time outlining what corpus stylistics entails and what its limitations are; and then I will share some ways in which I have tried to make it useful in the pursuit of my research interest, the textualization of narrative prospection.

Narrative prospection is itself only a stage in the experiential sequence of interest to me. I assume that the text's prospections cumulatively and serially guide the reader to expect the story currently being read to continue and terminate in one way rather than others (at the least, the prospections will foster probabilistic expectations). The ways in which the subsequent narrative text confirms or flouts

Michael Toolan is Professor of English Language in the Department of English, University of Birmingham, having previously taught at the National University of Singapore and the University of Washington, Seattle. His research interests and areas of publication include literary linguistics (e.g. *The Stylistics of Fiction*, 1988) narrative analysis (e.g. *Narrative: A Critical Linguistic Introduction*, 2nd edition, 2001), integrational linguistics (e.g. *Total Speech: An Integrational Linguistic Approach to Language*, 1996), and language in the legal process. He has books forthcoming on narrative progression, and on immersion and emotion in the reading of literary narrative.

these expectations, I hypothesize, are the bases of powerful cognitive-emotive responses, which are central to the reader's experience of the story. So in attempting to identify the textual sources of expectation, I am trying to work back to the textual roots of the reader's experiential immersion in a story. Equally, I am trying the better to model and understand reading a short story as a process and not merely as something leading to a product. The process is a phased one in which textual cues create expectations which in turn evoke feelings now of fear, now of suspense, now of surprise, menace, grief, anger, and so on, together with thoughts of unfairness, or futility, or disproportionality, or marvellous good fortune—the whole gamut of possible thoughts and emotions that fill our minds and absorb our attention in the course of our reading of an engaging narrative.[1] And the text, processed in sequence, is here regarded as the chief source, although not the only one, of the progression of emotional and cognitive responses that the reader has. Literary narratives are particularly powerful at causing the reader to feel immersed in a textual world, a world which seems increasingly interpretable and predictable as the text unfolds. This immersion is necessary for the accompanying emotional engagement that readers often allude to.

The methods I use here can be called corpus stylistic, but the latter is so nascent a subfield of corpus linguistics that the designation is presumptuous. Corpus linguistics itself, however, is well established. It involves the rapid searching and sorting of electronic versions of texts or language samples, often assessed comparatively against an appropriate reference corpus, i.e., a large computer-searchable gatherings of texts. Some versions of corpus analysis emphasize quantitative factors, and these can be important because of the unprecedentedly large data-sample on which the analyst's findings are based (Biber "Variation"; Biber et al.; Short and Semino; Hoover "Language"); 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 of style), with particular emphasis on the collocational properties of language (Sinclair "Corpus," "Trust"; Hunston and Francis; Louw "Irony"; Wray ; Stubbs "Words and Phrases," "Conrad,""Corpus analysis"; Hoey "Lexical Priming"). A growing number of scholars are now working on corpus-informed analyses of literary texts, and their work is of particular relevance to the present study (Carter; Hori; Herman; Yevseyev; Dillon "Corpus," "Genres"; Mahlberg ; Watson and Zyngier; Moon; and Hoover, Culpeper and Louw [forthcoming in 2008]).

Given an interest in prospection and expectation, can the rapid searching and sorting of word-forms and phrases enabled by corpus analysis, together with the theoretical commitments of corpus linguistics (collocation, phraseology, lexical priming), contribute to our understanding of the texture and structure of short stories? There are some grounds for optimism: some versions of corpus linguistic theory seem oriented towards the syntagmatic and the phrasal sequence (what 'normally' comes before and after a particular node word, for example—what Sinclair calls "the idiom principle" ["Trust"]). This is in contrast with standard language description, which focuses on the paradigmatic and on the grammatical categories rather than the lexico-semantic particulars which may or may not occur

(Sinclair calls this "the open choice principle"). Idiom analysts ask: what specific words tend to come before and after the target word? Open choice analysts ask: what syntactic category is the target word, and what are the categories from which preceding and following words must come? But it remains to be seen just how far 'open choice' corpus linguistics might reach, in its interest in the textual sequence. It explores a phrase like *naked eye* and shows that it is routinely (predictably) preceded by mention of some object that appears small to the unaided human perceiver, and then phrases such as *scarcely visible to,* or synonymous expression. Or the method can take a phrase such as *his hands slid [under/ over/ beneath . . .] her [blouse, breasts, elbows . . .]* and shows it is so frequent in romance fiction as to be a registral feature of that genre (Dillon "Genres"). But these are local spans of predictable wording, rather than long-distance predictions of the kind that the *narrative* stylistician craves. We seek them even as we realize, time and time again, that such long-distance expectation will and must be thwarted. Yes, we may guess, on the basis of a narrative's opening sentences about a prince being unsure whether any of his potential matrimonial partners is "a real princess," that by the end of the story someone that he is "sure" is "a real princess" will have been found or will have emerged; we might even guess that some test will be involved; but we cannot possibly predict, after the opening sentences, that a sleep-depriving pea will play the critical role that it does. The syntagmatic orientation of corpus linguistics in practice may be a quite local one, perhaps extending through simple clauses but without the kind of 'reach' necessary to be of assistance in narrative stylistic analysis. It is with such uncertainty that I proceed.[2]

Using text linguistic theory and corpus analytic methods in an attempted description of narrativity may sound somewhat abstract, but the issue is a fundamental one in the creation and understanding of text structure. "Oh what is going to happen now?" cries Bertha Young (at the end of Mansfield's story "Bliss"), at the anguished moment when she realizes her husband is having an affair with Pearl Fulton. But what is going to happen now or next—and later, and finally—are the sorts of questions that readers continually asks themselves in the course of reading any text. The read text prompts us to formulate expectations. Thereafter, those expectations inform a multitude of reactions and are hence crucial to the reading experience. Questions about what is going to happen next and later are only more sharply posed in relation to narrative text than to texts (and human behavior) in general. 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 sentences of a written narrative.

That expectation is a fundamental human propensity, invaluable to our ability to adapt to shifting circumstances, is generally recognized. But how we cope with 'the expected' relates directly to how we cope with the unexpected also, and as Chafe 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 each new day (this is a central theme in the integrational linguistic theory of which I am an adherent: see, e.g., Harris). As Chafe also notes, we even seem to welcome the unexpected, and the arousal (fear, anger, aggression) that it triggers, which in

turn helps us to cope with that unexpected: "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 83).

Textual progression has been an enduring interest of linguists at Birmingham University committed to a form-oriented discourse analysis, and is especially central to the work of the late John Sinclair, who argued that an accurate and powerful description of discourse needs to focus on the written or spoken text's prospective qualities. Sinclair combined an emphasis on prospection with a theory of encapsulation, in which each utterance in a text is assumed to provide a framework within which the next utterance is placed. Thus discourse is assumed to be designed in such a way that a first segment of text raises implicit questions that subsequent discourse will address, wholly or in part, and second and subsequent segments may open or prospect further kinds of question or incompleteness for yet later text to resolve (Sinclair "Trust," especially 82–101 and 115–127).

EXPECTATION

What can be said about the knowledge and the expectations that each successive word, sentence, and paragraph of a story causes readers to have or put in mind? In corpus linguistic analysis, the more mechanical or automatic ("analyst-neutral") the proposed mechanism of focussing or foregrounding is, the better, without embarrassment. My own procedures involve some crucial assumptions. First, I assume that while everything in a story is read, only a part of the text carries the main burden of signalling prospection or narrativity and by that means does the main work towards creating reader responses and expectations (suspense, surprise, tension, confusion or mystery, and so on). A second assumption is that this core narrativity-bearing material must, for psychological reality, be relatively easily noticed (that is, foregrounded text). Armed with these assumptions, and using ideas drawn from text linguistics, stylistics, and narratology, I seek to specify the most prominent materials—words, phrases, sentences—which (the argument runs) perform this core narrativity task, which therefore stand somewhat taller than the full circumambient text, and which, it is hoped, can be derived or pinpointed by relatively mechanical or automatic means. A short story is a complex texture, so a multiplicity of features or parameters (with no sharp cutoff point) is assumed to contribute to the conveying and apprehension of progression. I describe these briefly below, and discuss the prospects of being able rapidly to identify all and only their instantiations in a story in a principled and at least semi-automatic way. A major interim goal of the research is the deriving of a prospection-oriented abridgement of any story, a skeletal version of the story which contains the main triggers of readers' responses to perceived progression.

The words of James Michie's beautiful translation of lines from Horace's *carpe diem* ode (1.11) seem apposite here. Horace wrote: *Sapias, vina liques, et spatio brevi spem longam reseces*, which Michie renders as "Be wise, strain clear the wine / and prune the rambling vine / of expectation." We seem predisposed,

fired by emotions and imagination, to let the vine of expectation ramble extrava-gantly, even in the small disciplined arena of the short story. Pruning of the vine is necessary for long-term vigor and completion—flowering, fruiting, renewal—but this lesson is learned with experience, the realism of age. The short story can sim-ilarly feed the rambling vine of multiple and fantastic expectations, but the art of the writer lies partly in fostering some expectations rather than others, controlling readerly sympathies and antipathies, and causing us to be moved and to expect in particular productive ways, not all ways or any random way. That same *carpe diem* ode of Horace's figures importantly at the close of Alice Munro's story "Hateship, Friendship, Courtship, Loveship, Marriage", that story whose wilfully inelegant title, beginning with a word that does not exist (at least, not to be found in the OED online or my Random House desk dictionary), provides a crude Proppian plot summary. "Hateship, Friendship…" ends with our attention again focussed on the clever teenager Edith, whose cruel and irresponsible schemes have had enormous unforeseen *happy* consequences for others. Enid is doing her Latin translation homework, which just happens to be that same *carpe diem* ode from Horace. She has been tasked with construing the following instructive lines: *Tu ne quaesieris, scire nefas, quem mihi, quem tibi finem di dederint*: "You must not ask, it is forbid-den for us to know, what fate has in store for me, or for you—". But we *do* ask, we *do* want to know, what the text has in store for us as readers, and for its characters as our representatives and imagined others or doubles, in every literary narrative we read. And we are entitled to some guidance—not self-defeatingly comprehen-sive information, a pervasive prolepsis in which everything is told in reverse order—but some indications of how what we are told earlier connects with what comes later, even if it does not inescapably *cause* what comes later, and last.

There are, one must recognize, broad extratextual forms of textual guidance besides the guidance "in" the text. Among these are the story's particular genre or sub-genre and its norms of content and form, and the expectations that a particular topic creates, and that are further created by the story's context of situation—e.g., the assumed identity of the narrator and their role in events. Another kind of extra-textual guide may be our familiarity with the rest of the author's oeuvre, and the kind of characters she or he tends to portray (their gender, socioeconomic status, sexuality, age, job or profession, etc.). Consider the work lives of Alice Munro's main characters: they almost invariably work in moderate-status 'inconsequential' jobs, involving routine and repetition, and not high-prestige professions entailing the management or control of others, or the saving of lives, or the accumulation of wealth. A Munro story built around a powerful doctor or businesswoman would be deeply unexpected.

As for the *retention*, by the reader, of potentially guiding information, a plau-sible account of this area requires us to confront the huge verbal extent of the story narrative (to say nothing of the novel narrative) and the evidence, from introspec-tion, that much of what we read is and must be forgotten, to the point of being ir-recoverable even within a few lines. What the reader 'encapsulates' (to use Sinclair's term) and carries forward in active memory is normally no more than the gist of the gist. But I believe it is preferable (for reasons more fully set out in

Toolan, in preparation) to characterize what the reader takes forward, at any point in the reading of a story, as one or more maximally vague mental pictures with very few verbal labels attached. That picture (or pictures) is a visualization of the developing (and implicitly unstable or incomplete) situation that the text of the story so far has projected. That I believe is as much by way of 'mental representation' that a reader carries forward to the reading ('processing') of subsequent text—a maximally vague picturing accompanied by emotional, ethical and intellectual reactions and expectations. There is normally a necessary forgetting of the particulars of a long literary text (anything longer than an ode), without which those longer texts could not function. Milan Kundera has returned to this cruel paradox in his recent study, *The curtain: an essay in seven parts*, where he comments on the inevitable and warranted processes of forgetting that colour or structure every waking moment of our lives (not to mention when we are asleep; but it is as if we are mostly asleep even when awake): "the perpetual activity of forgetting." Resisting this forgetting, art strives to be treated as unforgettable; lyric poetry stands the best chance of being memorized and unforgotten, he says, while the novel by contrast "is a very poorly fortified castle." Nor does forgetting occur only between sessions of reading. Kundera writes: "[Forgetting] participates in the reading continuously, with never a moment's lapse; turning the page, I already forget what I just read; I retain only a kind of summary indispensable for understanding what is to follow, but all the details, the small observations, the admirable phrasings are already gone. Erased (Kundera 150). Thus cruelly abused, the novelist responds by striving to build "an indestructible castle of the unforgettable," by imposing form, composition, or what Kundera alternatively calls architecture. Composition is of exceptional importance to the novel—to all literary narratives, long or short, I interpret this as implying. For Kundera, composition is both distinctive to narrative verbal art and inseparable from its aesthetic claims: unlike other forms of art, "the beauty of a novel is inseparable from its architecture" (*ibid.*)

So an underlying assumption of an approach such as mine, which is so preoccupied with the linguistic detail of the narrative text, is that alongside the inevitable forgetting that perversely and paradoxically goes on as we attentively read a Munro story, there are forms of special attending, a metaphorical taking up or taking in, by the reader, of those textual articulations, hot spots, foregroundings, turning points, key evaluations, and so on, that the writer deploys so as to guide the reader's expectations as to the narrative's continuation, and their reactions at its end. But just as some architectural decisions are more important than others, so some composition choices are more important, more crucial and central than others, with multiple consequences for dependent compositional choices later. And yet we analysts do not actually know, for certain, what the most crucial or central choices are. Or, in terms of my interest in narrative progression, we do not know with certainty what the most crucial highlighted signals of a narrative's main themes and projected continuation are. Frankly we have to make informed hypotheses, working inductively, and then share them with other textlinguists and scholars of narrative and test the hypotheses out as best we can—sometimes not a very good best—on readers.

EIGHT TEXTUAL RESOURCES OF PROSPECTION

What kinds of 'waymarking' are particularly contributory to a foregrounding of prospection and expectation (cumulatively experienced and always responsive to the adjustments prompted by subsequent text) in the modern short story? I am proceeding on the assumption that a set of distinct but often overlapping textual features—distinct parameters of narrativity, in effect—tend to be most central to the creation of prospections and expectations. Currently I propose that eight such parameters are crucial:

Eight major textual resources ('parameters') in story prospection and the shaping of narrative expectation

1. Sentences in which occur 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).
6. Questions, requests/directives, undertakings and future-oriented informs, in Direct Speech.
7. Negation-carrying clauses: *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 actually clause-projecting (i.e., are followed by a full narrative proposition as complement: *he knew that it was no go, he knew that Corley would fail.*) The most important such verbs include the following: *know, think, seem, appear, suspect, expect, want, need, see, look, wonder, believe,* and *realize*.

These parameters or resources have been chosen on diverse grounds (explained in Toolan forthcoming). Some reflect the influence of authoritative work in discourse linguistics (from many sources, but especially ideas adapted from Longacre; Grimes; Halliday; Labov). Other parameters are included as they are recurringly identified as key material, prominent at the turning points and sites of crisis, in stylistic analyses (e.g. parameters 5 and 8) or narratological analyses (e.g. parameter 2) of the short story.

Parameters 1 and 4 concern what are known in corpus linguistic studies as keywords. The keywords of a text are a kind of "prominent repetition"; they are the disproportionately frequent word-forms in a text, when that text's word-frequencies are compared statistically with those in an appropriate reference corpus of texts (I prepared a small reference corpus—half a million words—of 20th century

British or American fiction, mostly short stories, mostly by women). The assumption is that the keywords (if there are any: it is quite possible for a text to contain *no* strikingly disproportionately frequent words) will be some pointer to a text's themes or preoccupations, and that by virtue of their recurrence they cannot easily be ignored by the reader. Parameter 4 focuses exclusively on prominent lexical or 'content' words. In "Two Gallants", for example, only *street(s)* and *walked* are lexical keywords; parameter 4 postulates that narrative sentences containing these words are foregrounded for prospection. In studies of reader responses to the opening paragraphs of Alice Munro's "The Love of a Good Woman" (Toolan 2008 and forthcoming), I tentatively concluded that a reader advances from the story's five-paragraph preface carrying forward (and expecting to learn more about) just two or three maximally vague images, with only a very few labels attached: W*illens, drown(ed), River, red box, instruments, escape, anonymous donor, optometrist's hand, black,* and *shiny*.

Some of these terms were found to relate directly to the automatically-calculable lexical keywords of the story's sections (keywords such as *water* and *river*) while other whole-story keywords are as it were unprepared for by the story preface (keywords such as *car* and *children*). If the emergence of *car* and *children* as keywords is unforetold, the reverse is true of *water* and *river*: because the opening paragraphs of "The Love of a Good Woman" appear particularly to prospect that the story will tell *inter alia* of how someone came to drown in a river, it is expectable that there will be prominent subsequent mention of water. Repeated use of the word *water* has a place already reserved for it in advance, given the preface's main prospections; as early as the close of the preface, *water* is foreseeable (and might even be called pre-read or pre-told).

Various kinds of adjustment and expansion of the lexical keyword "mapping" of story-texts were explored (and reported in Toolan forthcoming); one of the more interesting findings seems to be that such keywords tend to "bunch together," rather than being evenly distributed across the whole text; and this bunching or co-occurrence seems especially prominent in sentences crucial to prospection and narrativity. Despite these small advances, it is clear that using keyword analysis to probe narrative progression still leaves us a considerable distance from a sufficiently rich map of the main threads of narrative progression as reflected (assuming they are reflected) in lexical prominence and recurrence. The most obvious and immediate weakness relates to the difficulties that corpus analytic tools have with identifying what I call "para-repetitions": echoic (but non-identical) reiterations of incident or image using semantically similar but complexly-related lexical phrasing. In the Munro story, for example, references to the bashing and banging of heads recur prominently through the text (*braining . . . banging his head . . . head knocking against . . . hitting your head*), but without full lexical repetition and therefore in ways that evade capture by simple frequency calculation.

As for the first parameter listed above, this only emerged when I looked again at the keywords data for a variety of stories, and set aside my initial dismissive assumption about the comparative frequency of certain characters' proper names. I had first assumed that, in "The Dead" for example, of course *Gabriel* and *Gretta*

would be strikingly frequent "words" in comparison to their frequency in any suitable reference collection of texts, and that such prominences would therefore be uninteresting with regard to progression. Now suspending that dismissal, I focussed on the statistically topmost keyword in each story in a small sample, and found that the top keyword seems always to be the name—or part of the naming—of a prominent character in the story (but not the focalizing character). Thus in Carver's "Cathedral," the top keyword is *blind,* repeatedly used in the narration to name *the blind man* (or *this blind man*), whose given name is *Robert*; in Joyce's "Two Gallants" the top keyword is *Corley* (not that of the often-focalizing Lenehan); and in Updike's "A & P" it is *Lengel* (the supermarket manager).

Might there be after all, I wondered, a pattern or rationale behind this most-prominent name-use (often in places where a pronoun, or a variety of alternative descriptions, could have been selected)? If one takes all and only those sentences in the text where the top keyword features, a surprisingly coherent narrative text often emerges, or survives. Here, for example, is just a part of the "abridgement" of Carver's "Boxes" story, in which only those sentences containing the top character-naming keyword, *Jill*, have been retained:

> 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.

This is remarkable not only for being largely coherent, but also for what it manages to convey—something of that poignant, inarticulate, fumbling interpersonal connection between mother, son, and daughter-in-law (six relationships) that Carver captured in the original story. The abridgement I submit makes a surprising amount of sense, even to the reader unfamiliar with Carver's story—and it makes sense in ways that a textual abridgement around any other single word or keyword in the text seems not to. I have therefore tentatively proposed (Toolan "Keyword") that top character-designating keyword sentences are an important part of the "waymarking" of a narrative's progression; and that is why they are listed (as parameter 1) among the cluster of textual features which seem especially influential on the reader's moving picture (see below) of what is happening and what will likely happen in the story's continuation.

Each of the listed parameters, as will be clear, involves an explicit and even prominent part of the overt form of the text: it is rather important that the kinds of things that plausibly contribute to progression-expectations should be virtually unmissable, on the part of a reasonably attentive reader. And there is often overlap (which we might also regard as reinforcement, or overcoding) between the textual material that different parameters highlight. This is partly a consequence of my

decision to take the graphological sentence as the unit or text-span within which nearly all of these eight kinds of prospection foregrounding take effect. That, too, is a contentious but defensible claim—but not one that I can justify here. This privileging of the graphological sentence needs to be evaluated, and in the context of the methods espoused here, against alternative postulated units or segments (the clause? the graphological line? The paragraph? Or—e.g. for parameters 4 and 7—no more than the highlighted word or phrase itself?) in relation to which, it might be claimed, the identified feature has a promotional effect.

MECHANICAL FID-FINDING

A corollary to the noticeability requirement, directly relevant to corpus analytical searching, is that aspects of textual form that are necessarily noticed by every reader should also be amenable to at least semi-automatic identification. If a textual element is not so explicit and distinct as to be detectable and retrievable by automatic or rule-directed means, it is the less psychologically plausible that all reasonably careful readers will notice it; if we cannot be confident that they will notice it, we cannot claim it as a basis for the expectations they form. Noticeability is necessary (but not, of course, sufficient). Accordingly a significant part of the ongoing research effort is devoted to showing that the exponents of each of the proposed parameters can indeed be "captured" from a text by rapid mechanical (i.e. corpus linguistic) search. Relative to this goal, the "superficiality" of a parameter such as number 3, or even 7, is as much a strength as a weakness. But trying to reduce to an automatic procedure the textual moments described in parameters 2 or 5 presents more interesting challenges. Within parameter 5, for instance, what hopes have we of automating the retrieval of a discourse category as complex as Free Indirect Thought (part of parameter 5), traditionally regarded as resistant to context-insensitive identification? But then if the grounds for ascribing the presence of FIT cannot be rendered thus explicit, how confident can the analyst be that here is a phenomenon noticeable without difficulty by all competent readers, a site of prospection and expectation shifts that the normal reader can hardly miss?

Again my method of working on this problem involves shuttling inductively between reader-ratified evidence and simple text-based observations. No corpus linguist begins in the abstract, with "what FIT often looks like in literary narratives," but with specific putative instances. So I began with Joyce's "Two Gallants." Which of its sentences or sequences (or significant parts thereof) do readers and critics generally agree are FIT ones? The following five non-contiguous sequences are commonly recognized:

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.*

Finding no single pattern here, I propose multiple interacting criteria for the identification of FIT (in "Two Gallants" and, it is hoped, with application to other short stories at least of the Joycean-modernist kind). In effect, detection of this parameter is assumed to be multi-criterial just as the overarching textual mechanisms of prospection are assumed to be multi-factorial. Formulation of these criteria began with observing the prominence, in several of the reader-identified FIT sentences, of a main clause modal such as *would* or *might*, together with an immediately-adjacent personal pronoun (here, usually *he*). I propose three selection rules, to be applied in the order given here, as a means of identifying sentences that are good candidates for FIT designation:

1. In narrative (non-direct speech) sentences, select any that contain a modal verb (especially *would*, *could*, *might*, *must*, *should*) with Subject personal pronoun immediately before or after (*he*, *she*, *him*, *her*), where the construction is neither embedded under nor follows *if, whether, that, WH-*words*, nor other subordinators (hence modals within IT or IS are excluded)*, nor otherwise integrated into a complex grammatical structure. That is to say, modal plus pronoun should be in the matrix or main clause. A final condition is that where the pronoun precedes the modal, the latter must not be followed by *not* or *n't* (whereas with *never, nothing, nowhere* etc it remains FIT-indicative).
2. Select all sentences in the narrative (i.e. outside any direct speech) with in the main clause an interrogative structure or exclamative structure.
3. Select also any (non-direct speech) sentence, occurring immediately before and after one that is classified as putative FIT by rule 1 or 2, which contains any modal verb or modal adjunct (such as *could, would, should, must, might, surely, certainly, probably, possibly, maybe*), with or without negation.

The 'rules' as formulated here are still hybrids of functional and purely formal description, but my corpus-analytically-minded goal is to aim to express them in purely formal terms (thus, discarding opaque terms like 'embedded' and simply describing the search in terms of, e.g., *would* (etc.) with *he* (etc.) where *if, whether, that, why, how, when, where, who, which, because, since, while, unless, although, as soon as,* etc does not precede them without intervening major punctuation. And so on. To be sure, these criteria are imperfect; like all grammars (and boats), they "leak." But if they can be the basis (perhaps with author- or subgenre-specific modification) of capturing all and only the FIT sentences in a range of stories, to a high level of reliability, then I believe they merit a place as part of an equally leaky working model. We have to work with leaky models

where more watertight ones are currently unavailable (and perhaps even permanently and necessarily unavailable); otherwise we would never go to sea.

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.* (This sentence is caught by parameter 8, however: mental process verbs in the narrative that project a following proposition.) The three-stage procedure works reasonably well when applied to stories by Woolf, Kipling, Hemingway, and Joyce; it proved less successful on stories by Raymond Carver. Applied to Hemingway's "Cat in the Rain," for example, rule 1 at first selects just two possible candidates (sentences containing modal and pronoun):

1. *The cat was trying to make herself so compact that **she would** not be dripped on.*
2. *Perhaps **she could** go along to the eaves.*

But the first of these is immediately excluded on two grounds that are subconditions of rule 1: the *would* occurs in an embedded construction within a comparative (*so x that* PRO *would*); and direct-order pronoun plus modal is immediately followed by *not*. The second candidate, however, falls foul of none of rule 1's restrictions, so is automatically classified as FIT. And now rule 3 can be applied, and finds a modal in the textually-preceding sentence:

The cat would be around to the right.

This sentence is accordingly equally classified as FIT, yielding this sequence:

The cat would be around to the right. Perhaps she could go along to the eaves. There being no narrative sentences which meet the rule 2 condition, the above two sentences are the only FIT ones retrieved by the 3-rule procedure; they are the only sentences critics generally agree are FIT ones in the story.

It should also be acknowledged that the search procedures are incapable of distinguishing FIT from its reported speech counterpart, Free Indirect Speech (FIS). Accordingly it might have been more accurate to have referred above, throughout, to the umbrella 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

I have discussed the possibilities of 'mechanical' FIT recognition at length for two reasons. First, FIT is widely recognized to be important both to the plot and to character-development, in psychologically-oriented modern stories: it is highly significant to prospection and progression. Secondly, the three-stage search procedure suggests that FIT is more reliably detectable, via surface signals, than is sometimes supposed. That ease of detection supports the assumption that, con-

sciously or unwittingly, readers can hardly avoid attending, in the course of reading, to FIT segments and their deep disclosure of character preoccupations and expectations.

MAKING THE PICTURE MOVE

What have my corpus stylistic methods achieved? They have enabled me to visualize or reconfigure whatever story is currently under analysis in a distinct way, applying relatively robust, textlinguistically-warranted principles, and aiming for a high level of replicability. Anyone adopting the same list of categorizations and procedures as I have proposed should derive the same progression-foregrounding transform of an original story as I do. Whether the procedures I have nominated are the correct procedures, focussing on the most crucial textual components of progression, and whether some of these criteria or factors are much more important than others, remain matters for further assessment, by a variety of means. Specifying the narrativity features of a text and their importance relative to each other is difficult and controversial, but not in my view entirely impossible in principle.

What especially or particularly stimulates our interest, and our interest in pressing on to the end of the story (to see if things end the way we are coming to expect they will, or just to see how they end)? Could it be that text features or threads including the eight parameters proposed here would be central to the creation of narrative interest in the reader *because* they especially contribute to the reader's forming of vague but potentially-vividly-focused mental picturing of the ongoing changing Situation(s)? As intimated earlier, I conceive of the reader's processing of the passing narrative text as a forming of vague pictures that change: a show in which things in the pictures move. Everything in a narrative text has potential to contribute to this mental picturing (which is a fluid unresolved picturing not only of "what we are seeing" in/from the text, but also what we are hearing, tasting, smelling, feeling—potentially, all the senses, all the emotions, and all kinds of evaluative/ethical judgement). But some aspects of the picturing are more basic and shared, and basic to our willingness to read on to the end, than others (all readers will agree that Lenehan waits very anxiously for Corley; not all will agree that Lenehan is entirely to blame for his own sorry state). The prospecting, expectation-fostering parameters may be among the most basic resources for mental picturing. The narrative material instantiating these parameters, arguably, particularly causes the reader to make adjustments to their mental picturing (provisional and non-final adjustments) so that in reading such prospecting material we are especially aware that the Situation is not fixed, and that the mental picturing is a moving and changing one. The 8-parameter materials stimulate a heightened awareness (relative to the awareness fostered by other portions of the text) that the pictures (their contents) are moving. That awareness of movement is particularly acute at the "roller coaster" or "switchback" or "sudden reversal" moments in reading—these being terms in which readers frequently talk about narrative

surprises and narrative suspense. The sense or impression of movement (change) makes and keeps the narrative interesting to us, and draws us to read on until there is no more textual resource for making our picturings move.

It will be clear that all of the above is only a beginning, that an extensive application of these parameters on a range of stories is needed, to determine whether they indeed highlight those parts of the texts which, independent reader-response studies might help to confirm, are particularly instrumental in the shaping of the mid-reading expectations, and which in turn give rise to subsequent feelings such as suspense, surprise, uplift, anger, loss, and relief.

ENDNOTES

1. Literary narratives are only one source of such experiences. Some equally find such cognitive, ethical and emotional satisfactions in religious texts, or opera, or in the course of deep personal relationship with other individuals; in each of these examples the progressive or temporal dimension is crucial (the thoughts and emotions cannot all be experienced simultaneously but are necessarily encountered in sequence, just as the narrative text must be read in sequence, word by word or phase by phase).

2. One corpus linguist who seems to have no such doubts is Louw, who has speculated that a "Malinowskian revaluation" of stylistics should be possible when the full power of collocational analysis has been harnessed. 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: nnn)

WORKS CITED

Biber, Douglas. *Variation Across Speech and Writing*. Cambridge: Cambridge Univ. Press, 1988.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. *Longman Grammar of Spoken and Written English*. London: Longman/Pearson, 1999.

Carter, Ronald. *Language and Creativity: The art of common talk*. London and New York: Routledge, 2004.

Carver, Raymond. "Boxes." In *Where I'm Calling from: The Selected Stories*. New York: Vintage, 1989.

Chafe, Wallace L. "Some things that narratives tell us about the mind." In *Narrative Thought and Narrative Language*, edited by Bruce Britton and A.D. Pellegrini, 79–98. Lawrence: Erlbaum, 1990.

Dillon, George L. "Corpus, Creativity, Cliché: Where Statistics Meet Aesthetics." *Journal of Literary Semantics* 35:2 (2006): 97–103.

Dillon, George L. "The genres speak: Using large corpora to profile generic registers." *Journal of Literary Semantics* 36:2 (2007): 159–87.

Grimes, Joseph. *The thread of discourse*. The Hague: Mouton, 1975.

Halliday, Michael A.K. *An Introduction to Functional Grammar*. 2nd edition. London: Arnold, 1994

Harris, Roy. *Introduction to Integrational Linguistics*. Oxford: Pergamon, 1998.

Herman, David. "Quantitative methods in narratology: A Corpus-based Study of Motion Events in Stories." In *Narratology beyond literary criticism*, edited by Jan Christoph Meister, 125–49. Berlin: de Gruyter, 2005.

Hoey, Michael. *Patterns of lexis in text*. Oxford: Oxford Univ. Press, 1991.

Hoey, Michael. *Lexical priming: a new theory of words and language*. London: Routledge, 2005.

Hoover, David L. *Language and Style in <u>The Inheritors</u>*. Lanham, MD: Univ. Press of America, 1999.

Hoover, David L., Jonathan Culpeper, and William Louw. *Approaches to Corpus Stylistics: The corpus, the computer and the study of literature*. London: Routledge, forthcoming, 2008.

Hori, M. *Investigating Dickens' Style: A Collocational Analysis*. London: Palgrave, 2004.

Hunston, Susan and Gillian Francis. *Pattern grammar: a corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins, 2000.

Kundera, Milan. *The curtain: an essay in seven parts*. London: Faber & Faber, 2007.

Labov, William. *Language in the inner city: studies in the black English vernacular*. Philadelphia: Univ. of Pennsylvania Press, 1972.

Longacre, Robert E. *The Grammar of Discourse*. 2nd ed. (1983) London: Plenum, 1996.

Louw, William E. "Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies." In *Text and Technology,* edited by M. Baker et al. Amsterdam: John Benjamins, 1993.

Louw, William E. "Literary Worlds as Collocation." In *Literature and Stylistics for Language Learners: Theory and Practice*, edited by Greg Watson and Sonia Zyngier. London: Palgrave, 2006.

Louw, William E. "Collocation as the determinant of verbal art." In *Language and Verbal Art Revisited: Linguistic Approaches to the Study of Literature*, edited by D. R. Miller and M. Turci, 149–80. London: Equinox, 2007.

Mahlberg, Michaela. "Lexical cohesion: Corpus linguistic theory and its application in English language teaching." *International Journal of Corpus Linguistics,* 11:3 (2006): 363–83.

Moon, Rosamund. "Words, frequencies, and texts (particularly Conrad): A stratified approach." *Journal of Literary Semantics* 36 (2007): 1–34.

Munro, Alice. "The Love of a Good Woman." In *The Love of a Good Woman: Stories.* New York: Alfred Knopf, 1999.

Munro, Alice. "Hateship, Friendship, Courtship, Loveship, Marriage." In *Hateship, Friendship, Courtship, Loveship, Marriage*. London: Chatto & Windus, 2001.

Scott, Mike. *WordSmith Tools*. Oxford, Oxford Univ. Press, 1999.

Short, Mick, and Elena Semino. *Corpus Stylistics: Speech, Writing and Thought Presentation in a Corpus of English Narratives*. London: Routledge, 2003.

Sinclair, John M. *Corpus, Concordance, Collocation*, Oxford: Oxford Univ. Press, 1991.

Sinclair, John M. *Trust the Text: Language, Corpus and Discourse.* London: Routledge, 2004.

Stubbs, Michael. *Words and Phrases: Corpus Studies of Lexical Semantics.* Oxford: Blackwell, 2001.

Stubbs, Michael. "Computer-assisted text and corpus analysis: lexical cohesion and communicative competence." In *The Handbook of Discourse Analysis*, edited by D. Schiffrin, D. Tannen, and H.E. Hamilton, 304 – 320. Oxford: Blackwell, 2001.

Stubbs, Michael. "Conrad in the computer: examples of quantitative stylistic methods." *Language and Literature* 1:5 (2005) 5–24.

Stubbs, Michael. "Corpus analysis: the state of the art and three types of unanswered questions." In *System and Corpus*, edited by Susan Hunston and Geoff Thompson. London: Equinox, 2006.

Toolan, Michael. "Top keyword abridgements of short stories: a corpus linguistic resource?" *Journal of Literary Semantics* 35:2 (2006): 181–94.

Toolan, Michael. "The Language of Guidance." In *Theorizing Narrativity*, edited by John Pier and José Ángel García Landa. Berlin: de Gruyter, 2008.

Toolan, Michael. (forthcoming). *Narrative Progression in the Short Story: a corpus stylistic approach*. London: Palgrave, 2008.

Toolan, Michael. *Immersion and emotion: how textual patternings shape our experiencing of literary narrative.* (In preparation).

Watson, Greg and Sonia Zyngier, eds. *Literature and Stylistics for Language Learners: Theory and Practice*. London: Palgrave, 2006.

Wray, Alison. *Formulaic language and the Lexicon*. Cambridge: Cambridge Univ. Press, 2002.

Yevseyev, V. "Measuring narrativity in literary texts." In *Narratology beyond literary criticism: mediality, disciplinarity*, edited by Jan Christoph Meister, 109–24. Berlin: Walter De Gruyter, 2004.