

Pamphlet **5**
→ June 2013

Literary **Lab**

Style at the Scale of the Sentence

Sarah Allison

Marissa Gemma

Ryan Heuser

Franco Moretti

Amir Tevel

Irena Yamboliev

Pamphlets of the Stanford Literary Lab

ISSN 2164-1757 (online version)

Sarah Allison
Marissa Gemma
Ryan Heuser
Franco Moretti
Amir Tevel
Irena Yamboliev

Style at the Scale of the Sentence

1. A Discussion on Style [April 2011]

In April 2011, the Literary Lab held a broad retrospective discussion of its first year's work. Among other things, we talked about our first pamphlet—"Quantitative Formalism"—and wondered whether its object had indeed been style, as we had rather casually claimed at some points in the pamphlet, or whether the method followed had been too reductionist to capture that elusive object. A chart from "Quantitative Formalism" will clarify the point: in [Figure 1.1](#), Jacobin novels (on the right of the diagram) are separated from Gothic ones (on the left) by the different frequency of the words that appear in the chart: "you", "if", "not", "she", "were", "the" and all the other terms included in a program created by Matt Jockers, which we ended up calling Most Frequent Words, or MFW. For Gothic novels, distinctive MFW included third person pronouns, verbs in the past tense, locative prepositions, articles, and more; for Jacobin ones, second person pronouns, a conjunction expressing uncertainty ("if"), and the conditional auxiliary "would".

The units employed by MFW (articles, pronouns, prepositions, etc.) were clearly functional to the central aim of the two forms: in fact, it was precisely *because* they were so profoundly functional to narrative suspense (for the Gothic) and argumentation (Jacobin novel) that they were so good at separating the two genres. But could the different frequencies of "she" and "you" and "the" really be called "style"? On this, we disagreed. Some of us claimed that, though all styles do indeed entail linguistic choices, not all linguistic choices create style; others countered this argument by stating that style follows necessarily from this fundamental level, and that all we need to analyze it is the set of linguistic choices made by an author or genre. This was the genuinely reductionist position—style as *nothing but* its components—and the more logically consistent one; the other position admitted that it couldn't specify the exact difference, or the precise moment when a "linguistic choice" turned into a "style", but it insisted nonetheless that reducing style to a strictly functional dimension missed the very point of the concept, which lay in its capacity to hint, however hazily, at something that went *beyond* functionality. Our job should consist in removing the haze, not in disregarding the hint.

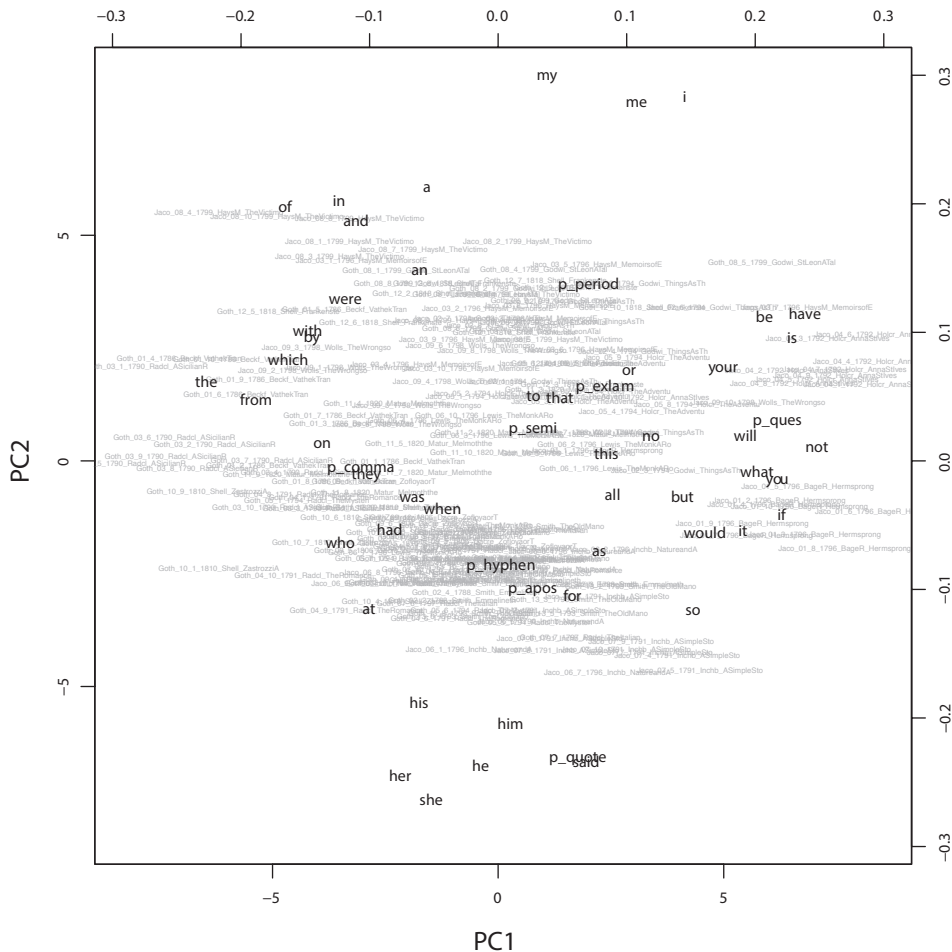


Figure 1.1: Most Frequent Word scatterplot with titles (light grey) and component loadings (black). From Literary Lab Pamphlet 1, “Quantitative Formalism: An Experiment”.

We will return at the end to the “not merely functional” nature of style. For now, let’s just say that, since the anti-reductionist position was the more numerous one, we used it as the basis for developing the next stage of the argument, by considering a series of linguistic structures of increasing complexity to try and capture the moment at which style became visible. The series went something like this: Gothic novels have many locative prepositions; but a thousand occurrences of “from”, “on”, “in”, and “at” are not style in any conceivable sense of the word. Jacobin novels have a lot of conditionals; a little better, perhaps, but not much. Then came the formula Franco Moretti had noticed in Gothic titles, and analyzed a few years earlier in “Style, Inc.”: “the x of y”: *The Castle of Otranto*, or *The Rock of Glotzden*. The formula was a perfect expression of the Gothic obsession with space; but, once more, functionality was not really style. The next layer was another formula, that Marissa Gemma had identified in Poe, and discussed in her dissertation: “the x of y of z”, she had called it—as in *The Fall of the House of Usher*, or “the gray stones of the home of his forefathers”. This authorial exaggeration of a generic trait, with its defiance of any mere functionality, offered a first glimpse of what we were looking for; maybe it was style, maybe it wasn’t, but we were finally getting close. And with the next instance—the opening words of *Middlemarch*: “Miss Broke had that kind of beauty which seems to

be thrown into relief by poor dress”—we all agreed we had entered the territory of style proper. As Sarah Allison had shown in her dissertation’s analysis of this type of sentence, a whole series of connections and transformations coalesce around the relative pronoun “which”: as the past tense of the main clause becomes the present of the dependent one, narrative distance turns into engaged comment, and character description (“Miss Brooke had beauty”), into a nuanced qualification of the type and meaning of that beauty.¹ One reads the sentence, and immediately gets the sense of a work capable of modulating from novel into essay, and from the relative simplicity of the story to the subtlety of reflection. The sentence is certainly perfectly functional to the opening of a novel—but it also possesses many other layers of meaning, all closely interconnected. Now, *this* was style.

We had found a starting point. We would study not style as such, but style *at the scale of the sentence*: the lowest level, it seemed, at which style as a distinct phenomenon became visible. Implicitly, we were defining style as a combination of smaller linguistic units, which made it, in consequence, particularly sensitive to changes in scale—from words to clauses to whole sentences. Yet we also hesitated, because the sentence wasn’t at all an obvious choice for stylistic analysis; Auerbach in *Mimesis*, or Watt in his essay on *The Ambassadors*, had, for instance, operated at the quite different scale of the *paragraph*: ten, twenty, thirty lines, that included a much greater variety of linguistic traits, and could thus be seen (most clearly in *Mimesis*) as a model and miniature of the work as a whole. Sentences seemed much too short to play the same role. Perhaps they could play a different one? Did something happen at the scale of the sentence *that could not happen at any other scale?*

2. Sentence Types: the Initial Choice [October 2011-January 2012]

We would be studying sentences, then. And, given that a long tradition of narrative theory—from Benveniste’s *Essays in General Linguistics* to Barthes’ *Writing Degree Zero* and Weinrich’s *Tempus*—had recognized a categorical difference between story and discourse, we began by separating the sentences that belonged to the dialogue among characters from those that belonged to the narrative system. We needed texts where speech was marked with enough clarity and consistency for our tagger to recognize it, so we turned to the Chadwyck-Healey nineteenth-century database (smaller than our Lab corpus, but cleaner, with about 250 well-marked British novels), and separated the sentences into three types: those containing dialogue, those containing a mixture of narrative and dialogue, and those containing only narrative. Allison and Moretti concentrated on the “mixed sentences”, where the intersection of dialogue and narrative—which had not been much studied by narrative theory—seemed to promise interesting stylistic effects; but this line of inquiry quickly became so specific that we decided it would require a study of its own. Meanwhile, Gemma, Heuser, Tevel, and Yamboliev chose to focus exclusively on narrative sentences, and on a few well-defined combinations of clauses. **Figure 2.1** shows the one- and two-clause sentences that quickly emerged as the most prevalent in the corpus.

¹ A version of this argument is forthcoming in *ELH* as “Discerning Syntax: George Eliot’s Relative Clauses”.

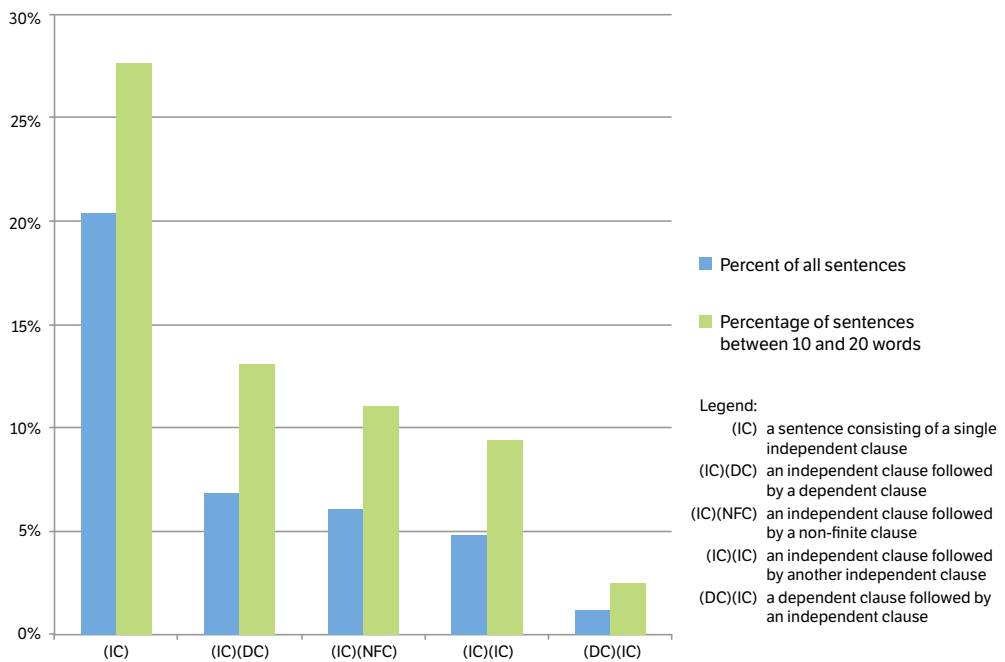


Figure 2.1: Distribution of sentence types by clause combination in the nineteenth-century novel corpus. As the chart shows, these five sentence types account for 40% of all narrative sentences, and about 65% of sentences between 10 and 20 words.

Initially, the group concentrated on three types above all others: IC-IC sentences, which consisted of two independent clauses; IC-DC, where an independent clause was followed by a dependent one; and DC-IC, where the dependent clause preceded the independent one. Two-clause sequences established a relationship between propositions that—in line with our initial plan—might allow us to see style emerging from their combination, while remaining small enough to capture the narrative or semantic logic contained therein. And at this point, the turning point of the entire project occurred. Allison suggested that we start with conjunctions, as they provided a grammatical condensation of a logico-semantic relationship—adversative, causal, coordinating, correlative, defining, predicative, etc.²—and were thus the perfect place to begin our investigation. By and large, we expected that the *distribution* of the logico-semantic relationship would be extremely variable—some texts inclining towards the causal register, or the predicative, or the coordinating one—while the *order* of the clauses would often be completely unrelated to logical function; so that, for example, a text with a preference for “narrative sequencing”³ would be equally likely to express such relations in IC-DC sentences as in DC-IC ones.⁴ But the results of our inquiry—summarized in [Figure 2.2](#)—proved to be quite different from our expectations.

For us, the most striking aspect of this figure was the radical asymmetry between two logico-semantic relationships, and two sentence-types: the “sequencing” relation, that appeared in 51% of the DC-IC sentences, but only in 13% of the IC-DC ones, and the “defining” relation, that appeared in 41% IC-DC, and a mere 5% of DC-IC. The asymmetry was so marked that, at the meeting when it was first presented, it was received with a lot of ambivalence: though the IC-DC findings could be explained by grammatical necessity, what about the DC-ICs? We had been looking for the emergence of style (“Miss Brooke had that kind of beauty...”)—but the structure of DC-IC sentences seemed to alert the reader to *narrative* developments instead. From the very first word, its inner form implied a preparation, then a pause—“When the day came round for my return to the scene of the deed of violence,”—and then, after the comma, the rapid completion of the mini-sequence (“... my terrors reached their height”: Dickens). When, as in [Figure 2.3](#), we noticed that 88% of Radcliffe’s DC-IC sentences had a “sequencing” function, we felt we had found

2 Needless to say, grammatical nomenclature nowadays is highly variable: in the main, we have followed the categories of Rodney Huddleston and George K. Pullum, *The Cambridge Grammar of the English Language*, Cambridge U.P., 2002, pp. 1293-1321.

3 Our notion of a “sequencing” relation between clauses is based on *The Cambridge Grammar’s* discussion of “temporal sequence”, (Huddleston and Pullum 1300) which—as we aimed to capture the full spectrum of temporal ordering conveyed by conjunctions we expanded to include both coordination and subordination. “Sequencing” relations are for us those in which the conjunction creates any temporal order—linear, non-linear, and simultaneous. As such, sentences like “Before the shades of evening had closed around us, I had a dozen awakening letters for my aunt, instead of a dozen awakening books” (Collins), and “While I was anticipating the terrors of a heroine, he introduced me to his Cardinal” (Disraeli), are also tagged as “sequencing” sentences.

4 On the other hand, we did *not* expect the order of clauses to be unrelated to function in the case of the “defining” relation. Like the category of “sequencing”, we based our category of “defining” relations on grammatical terminology: a defining sentence is one in which the relative clause— i.e., a dependent clause using “which”, “who”, or “that”— defines or characterizes the other clause: “This was Mrs. Finn, the wife of Phineas Finn, *who had been one of the Duke’s colleagues when in office*” (Trollope). Since it is nearly impossible to place a dependent defining clause before the clause that it elaborates, it makes grammatical sense that we would find more defining sentences in IC-DCs (like the Trollope example here). And indeed, our findings for the order of clauses among “defining” sentences conformed neatly to the demands of grammatical correctness—and thus proved rather unproductive for our literary analysis. This finding was also confirmed by our semantic analysis of IC-DC sentences, as we discuss in section 4.

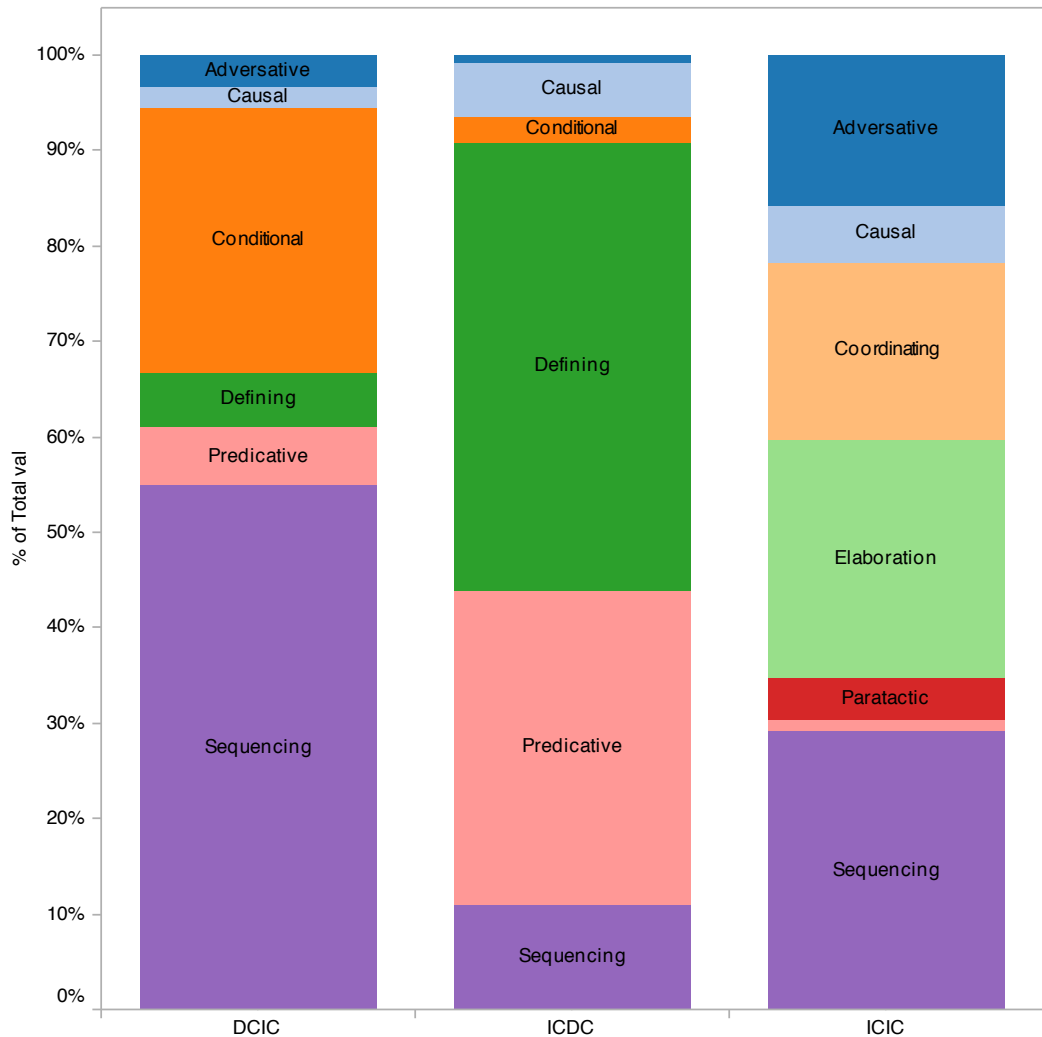


Figure 2.2: Distribution of the prominent clause relations across sentence type. Notice the almost perfectly inverse relationship between the DC-IC, dominated by “sequencing” and “conditional” relationships, and the IC-DC, almost entirely taken by “predicative” and “defining” ones.⁵ The IC-IC, for its part, resembles neither of the other two sentence types, and specializes in coordination and, to a lesser extent, parataxis.

a very significant metric of narrativity— especially when we compared these results with our non-narrative control text, Darwin’s *The Origin of Species*, in which such sentences were nearly absent (2%). But this metric seemed to have little to do with the concept of style. And when, a few weeks later, Amir Tevel found some unexpected narrative traits in IC-DC sentences, the switch in focus from style to narrative seemed even more inevitable.

⁵ We used “defining” to categorize both restrictive and non-restrictive relative clauses, since we found that distinguishing between them did not substantively add to our analysis. For the fine-grained distinctions between these types of clauses, see Huddleston and Pullum, 1033-1064.

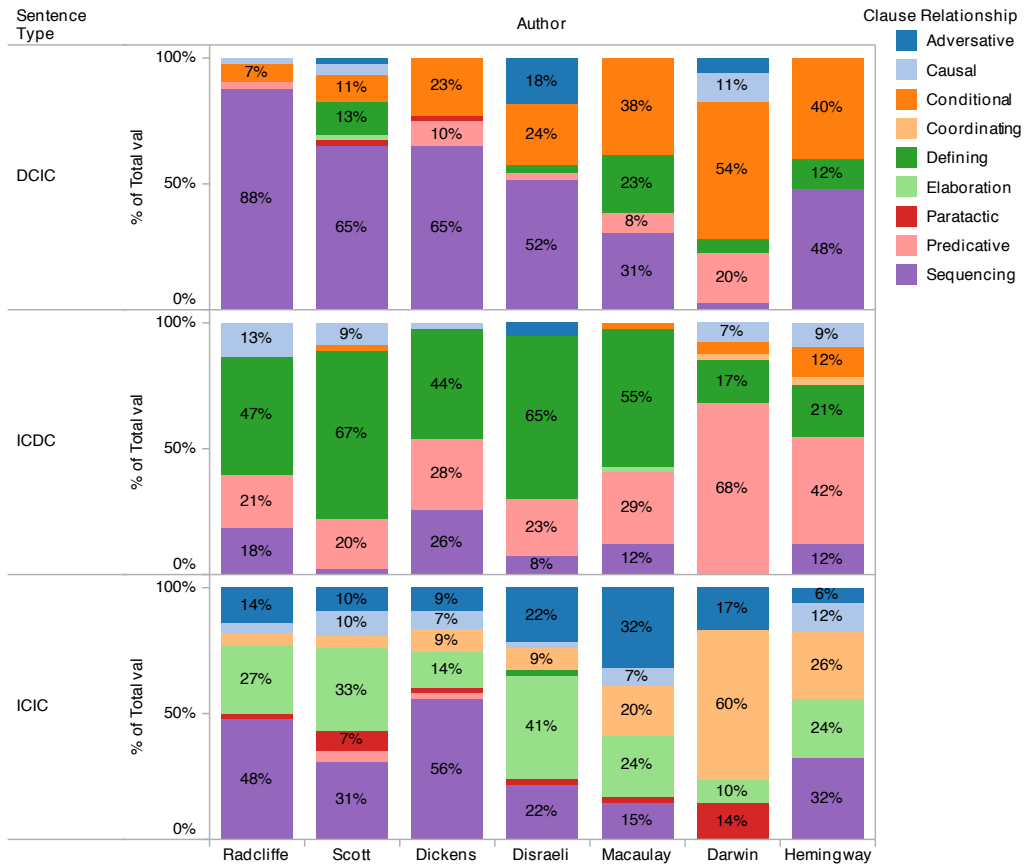


Figure 2.3: Sentence types: the spectrum of possibilities. Some of the authors in our corpus (like Dickens and Radcliffe) overwhelmingly favored sequencing over other possible relations, while others (Scott and Disraeli) were more moderate, and, most strikingly, Darwin used almost no sequencing (2%), compared to Radcliffe's stunning 88%.

3. Towards a Typology of Narrative Sentences [February 2012]

While working on the structure of IC-DC sentences—which, as we have noted, generally have much more to do with predication and definition than with sequencing—Tevel noticed among predicative and defining sentences an embedded narrative configuration that seemed typical of them. Here are a few examples:

3.a

Her extreme beauty softened the inquisitor who had spoken last. [P. B. Shelley]

But no matter; I will be the friend, the brother, the protector of the girl who has thrown herself into my arms. [Dacre]

It was then offered to the Palmer, who, after a low obeisance, tasted a few drops. [Scott]

He uttered an involuntary exclamation, and called to the driver, who brought the horses to a stop with all speed. [Dickens]

Fanny called the post-boy to the window of the chaise, and gave him directions, at which he a little stared, but said nothing. [M. Shelley]

In looking at these sentences, Tevel noticed that the dependent clause did two things at once: it introduced a *different character* from the subject of the main clause—the post-boy, the driver, the inquisitor, the Palmer, the girl who had thrown herself—while also allowing these newcomers a *very limited role* in the text: the post-boy stares but says nothing, the driver stops the horses, the Palmer tastes a few drops. It's an opening of the story to the Many—to use Alex Woloch's term for minor characters—but these Many get to do only a Little. A little, in the sense that they complete an already-initiated sequence rather than inaugurate an independent action. The syntax itself nudges writers in this direction: since it's hard to imagine a dependent clause that does something independent from the main clause, these (half-)sentences slide almost “naturally” into a form of *narrative attenuation*. They narrate—but minor episodes only. Conan Doyle used the unconscious expectations arising from this grammatical fact to perfection when he placed clues in dependent clauses, thus making them visible while suggesting to readers that nothing important was being said.⁶ In the following example, for instance, the smell of the cigar, which is the decisive piece of information, appears only as the third link of a tight chain of subordinates, and is further deflected in a non-narrative direction by the relative clause that follows:

That fatal night Dr Roylott had gone to his room early, though we knew that he had not retired to rest, for my sister was troubled by the smell of the strong Indian cigars which it was his custom to smoke. (*The Adventure of the Speckled Band*)

This, then, was the characteristic narrative function of these IC-DC sentences. And when we turned to DC-IC sentences, the mirror-image configuration emerged: here, the depen-

⁶ As always, Viktor Shklovsky understood it all a century ago: “[In the Holmes stories] instructions are given not directly but in passing (i.e., in subordinate clauses, on which the storyteller does not dwell, but which are nonetheless of major importance) [...] the clue is intentionally placed in the oblique form of a subordinate clause”. Viktor Shklovsky, *Theory of Prose*, 1929, Dalkey Press, Champaign, IL 1991, p. 106.

dent clause—which of course in this case *preceded* the main clause rather than following it—tended to report a muted preparatory event, while the main clause included the more surprising one:

3.b

While she looked on him, his features changed and seemed convulsed in the agonies of death. [Radcliffe]

When it was once fairly put before her, the effect was appalling. [Disraeli]

When she awoke, it was to the sound of guns. [Eliot]

In all these examples, the shift in grammatical subject—she/his features; it/the effect; she/it—coincided with an increase in narrative intensity: a semantic *crescendo*—agonies, appalling, explosions—that mirrored the *diminuendo* we had found in IC-DC. It was as if these two types of sentences embodied the systole and diastole of the narrative system: contraction-attenuation in IC-DC, and expansion-intensification in DC-IC. As both functions are indispensable to story-telling, we decided to try and find out whether expansion and contraction alternated as regularly in novels as they do in living organisms, and we began by looking for the diastolic-systolic patterns of the other three most frequent types of narrative sentences (IC, IC-IC, IC-NFC). Here, the most interesting result was Yamboliev's discovery that, in a (relatively small) group of IC-IC sentences, the relationship between the two clauses was one of slight elaboration, or reiteration, or restatement: in other words, fundamentally, of *stasis*:

3.c

Perseverance alone was requisite, and I could persevere. [Holcroft]

She raised her head; she lifted her hand and pointed steadily to the envelopes. [Collins]

Oh she looked very pretty, she looked very, very pretty! [Dickens]

Will Ladislaw, meanwhile, was mortified, and knew the reason of it clearly enough. [Eliot]

He showed no sign of displeasure; he hardly noticed. [Barry]

As stories have to intensify, attenuate, and remain in some way static, Yamboliev's finding seemed intuitively right; and when a large group of IC-NFC, with gerunds in the dependent clause, added the nuance of actions moving in parallel to each other, and overlapping in the process, a genuine typology of narrative sentences seemed within reach—and with it, the possibility of “sequencing” entire novels, charting the distribution of narrative intensity throughout their length.

But there were two obstacles on the road to the narrative genomics we were beginning to envision. First, in order to identify the signs of narrative intensity in the thousands of sentences contained in a single novel, or the millions of a broader corpus, we had to find a way to machine-gather the evidence. Our parser was however far from perfect even at recognizing the five main sentence types; anything needing a finer grain—like expansion, contraction, inertia and so on—would make it completely unreliable. And then, the

two-clause sentences with a “sequencing” function that we had singled out were a small minority of a novel’s narrative sentences;⁷ even smaller, if dialogue and “mixed” sentences were restored to the mix. If we wanted to chart narrative rhythm *at the scale of the novel as a whole*, we had to find a way of systematically integrating our small subset of sentences into the rest of the novelistic system.

We realized that ambition required patience. A general typology of narrative sentences, so clearly defined as to be recognizable by a computer program, needed its own *ad hoc* project.

4. Semantic Vectors [March-April 2012]

What we had found so far were striking correlations between particular syntactic forms and two seemingly unrelated domains—logical relations like “causal” or “sequencing”, and effects of narrative rhythm like “attenuation” or “stasis”. How far might these links extend? Could sentence structures be so powerful as to make certain words more likely to appear, thus establishing a link between syntax and semantics? This additional correlation seemed unlikely, but worth exploring: in contrast to higher-order domains like logic and narrative rhythm (which require human annotation), the likelihood of a word’s occurrence can be quantified automatically by a computer. We took all the words in all the sentences of our corpus, and calculated their average, or “expected”, frequency; then we calculated the actual—or “observed”—occurrence in each sentence type; finally, we focused on those instances that revealed a significant observed-over-expected ratio. We had developed this procedure, called “Most Distinctive Words”, in a study of drama (the subject of one of our pamphlets). **Figures 4.1** and **4.2** give an idea of the results, showing the “distinctiveness” of the most frequent words (**4.1**), and verbs (**4.2**), in the four sentence types.

⁷ If only taking into account two-clause sentences, the total of these “sequencing” sentences would be about 9% of all narrative sentences; if including also single-clause ICs (those that “sequence” a shift between states or events in a single clause) the total would rise to 28%.

Observed over Expected Values for the Top 100 Words in IC, ICDC, ICIC, DCIC

word	document			
	(IC)	(IC)(DC)	(IC)(IC)	(DC)(IC)
the	1.0	1.0	1.0	0.9
and	1.0	0.7	1.4	0.7
of	1.1	1.0	0.9	0.9
a	1.1	0.9	1.0	0.8
was	1.0	1.0	1.1	0.9
in.	1.0	1.0	1.0	0.9
he	0.9	1.2	1.0	1.4
his	1.0	0.9	1.0	1.0
to	1.0	1.0	0.9	1.0
her	1.0	0.9	1.1	1.0
had	0.8	1.3	1.0	1.2
with	1.1	0.9	0.9	0.9
it	0.9	1.2	1.1	1.0
she	0.9	1.1	1.0	1.3
at	1.1	0.9	0.9	0.8
that	0.5	2.4	0.4	0.9
i	0.9	1.2	1.0	1.3
but	1.0	0.6	1.6	0.8
not	1.0	1.0	1.0	1.0
for.	1.0	1.0	1.0	0.9
on	1.1	0.9	1.0	1.0
as	0.7	1.5	0.7	2.3
s	1.1	0.9	0.9	0.9
him	1.0	1.0	1.0	1.1
were	0.9	0.9	1.3	1.2
this	1.1	0.9	0.7	0.8
by	1.0	1.0	0.9	0.9
from	1.0	1.1	0.9	0.9
all	1.1	0.9	1.0	0.8
my	1.1	0.9	1.0	1.0
they	0.8	1.0	1.3	1.6
there	0.9	1.0	1.1	0.8
been	0.8	1.2	1.1	1.3
no	1.1	0.9	1.1	0.7
is	1.0	1.0	0.9	0.8

(continued)

word	document			
	(IC)	(IC)(DC)	(IC)(IC)	(DC)(IC)
mr	1.1	0.9	0.9	0.9
an	1.0	1.0	1.0	0.8
one	1.0	1.1	1.0	0.8
have	0.8	1.2	1.0	1.7
would	0.8	1.2	1.0	1.7
could	0.8	1.3	0.9	1.5
their	1.0	0.9	1.1	1.0
which	0.1	3.4	0.1	0.4
more	1.0	1.0	1.0	1.4
be	0.8	1.4	0.9	1.5
so	0.9	1.4	0.8	1.0
or	1.0	0.9	1.1	1.1
very	1.1	0.9	1.0	0.6
up	1.1	0.9	1.0	1.2
them	0.9	0.9	1.2	1.2
then	1.1	0.7	1.2	0.3
now	1.1	0.9	1.1	0.7
into	1.0	1.0	1.0	1.2
than	0.9	1.3	0.8	0.7
little	1.1	0.9	1.0	0.9
out	1.0	0.9	1.0	1.1
me	1.0	1.1	1.0	1.0
did	0.9	1.1	1.0	1.2
time	1.0	1.0	1.1	0.7
upon	1.0	1.0	0.9	1.2
some	1.0	1.1	1.0	0.8
man	1.0	1.0	0.9	0.9
mrs	1.1	0.9	0.9	1.0
we	0.8	1.2	1.0	1.6
when	0.1	2.0	0.1	9.2
two	1.1	0.9	1.1	0.8
other	1.0	0.9	1.0	0.9
looked	1.2	0.8	0.9	1.1
came	1.0	0.9	1.0	2.3
lady	1.1	1.0	0.9	0.8

Figure 4.1: Sentence types and their most distinctive words. Dark green indicates a strongly above-average frequency, and dark red a strongly below-average one; grey indicates a frequency close to the average. Some results literally leap to the eye: in 4.1, for instance, “which” is virtually absent from IC and IC-IC, very rare in DC-IC, and extremely frequent in IC-DC, where it introduces dependent relative clauses;⁸ while “when” occurs 9.2 times (!) above its expected rate in DC-IC. Same for the verbs of Figure 4.2, where “came” is particularly frequent in DC-IC, “looked” and “took” in IC, “knew”, “felt”, and “thought” in IC-DC.

⁸ These findings for IC-DC sentences—unsurprising in themselves—help establish the accuracy of our semantic results, as, grammatically, one would expect to see “which” in sentence types likely to have a defining or predicative function, as in “Five or six rings and a bracelet had been taken also from Lizzie’s dressing-case, which she had left open” (Trollope).

Observed over Expected Values for the
Top 100 Verbs in IC, ICDC, ICIC, DCIC

word	document			
	(IC)	(IC)(DC)	(IC)(IC)	(DC)(IC)
was	1.1	1.0	1.1	0.8
had	0.8	1.1	1.0	1.0
were	0.9	0.8	1.2	1.0
been	0.7	1.1	1.0	1.0
have	0.8	0.9	0.8	1.3
be	0.7	1.2	0.7	1.2
is	1.3	1.0	1.0	0.8
did	1.0	0.9	1.1	1.1
came	0.9	0.6	0.8	1.7
went	1.3	0.7	0.9	1.2
made	1.0	1.3	0.9	0.8
looked	1.5	0.7	1.0	1.0
are	1.1	0.9	1.2	0.9
took	1.6	0.8	1.1	0.8
knew	0.7	1.6	0.9	0.7
said	0.8	1.1	0.8	1.2
found	1.1	0.9	0.9	1.1
left	1.0	0.8	0.9	1.3
heard	0.7	1.1	0.9	1.2
has	1.2	1.2	0.9	0.8
saw	1.0	1.1	0.8	1.0
turned	1.4	0.7	1.0	1.1
felt	1.0	1.6	0.7	0.7
passed	0.9	1.0	0.8	1.2
thought	0.5	1.6	0.7	1.0
done	0.5	0.8	0.7	1.7
stood	1.2	1.1	0.9	0.9
do	0.7	1.4	0.5	1.3
gave	1.2	0.9	1.3	0.8
sat	1.2	0.9	1.1	0.9
became	0.9	0.7	1.0	1.3
put	1.2	0.9	0.9	1.1
entered	1.1	0.7	0.8	1.4

Figure 4.2: Sentence types and their most distinctive verbs. See Figure 4.1 for explanation.

Now, this was clearly a good way to recognize the role of individual words in different sentence types; fortuitously, small semantic clusters also emerged—like the “knew-felt-thought” group in IC-DC, which suggested a focus on knowledge and perception. But the approach remained fundamentally atomistic: analysis could only proceed one word at a time. By contrast, principal component analysis—which we used in “Quantitative Formalism”—offered a synthetic view of the entire semantic distribution of the four sentence types with a single image (Figures 4.3 and 4.4):



Figure 4.3: Sentence types and their semantic space (all words)

These charts use exactly the same data as 4.1 and 4.2, but visualize them differently, expressing several variables at once. Font size indicates the “distinctiveness”, that is to say, the observed-over-expected ratio of a word: the rather large “having”, at the top of Figure 4.4, occurs 2.4 times more frequently than one would expect it, whereas the slightly smaller “came”, on the lower right, only 1.6 times. Color indicates for its part the absolute value of occurrences: the gold of “came” tells us that the word occurs between 1,000 and 10,000 times in the corpus, whereas “having”, in green, occurs only between 100 and 1,000 times: though *less distinctive* than “having”, “came” is therefore *more frequent* in the corpus.

In addition to font size and color—that is to say, relative and absolute frequency—what most matters here is the spatial position of the various words: “came” on the right side, and “having” at the top, are where they are because they are correlated with the red lines (the “vectors”) that represent, respectively, DC-IC and IC-IC; while “knew”, “felt”, and “thought” are on the left side because they are correlated with IC-DC. The findings of Figures 4.1 and 4.2 are here visualized in such a way that the differentiation between the four sentence types is expressed by several words at once. What also emerges more clearly than in the previous figures is that—although there are four sentence types involved—there are only *three* semantic “positions”: IC and IC-IC tend to share the same semantic traits, and hence the same space in the diagram. This suggests, somewhat surprisingly, that, whereas the introduction of a dependent clause produces a distinct *semantic space*, the introduction of a second independent clause is characterized mostly by redundancy and repetition.

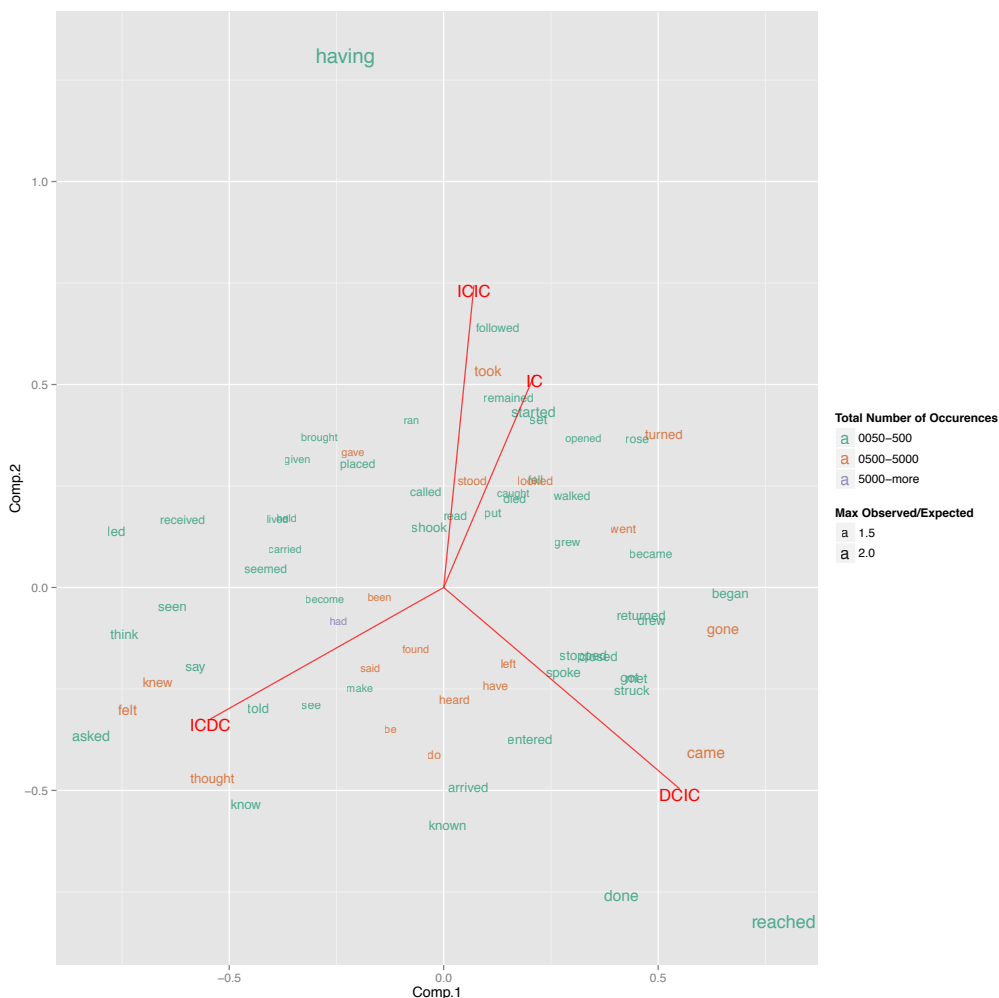


Figure 4.4: Sentence types and their semantic space (verbs)

Looking at the semantic differentiation of [Figures 4.3 and 4.4](#), the differences between the four sentence types—or three, considering that IC and IC-IC occupied the same semantic space—were acquiring more definite contours. In several cases, though, the results were somewhat puzzling: finding “home”, “door”, and “change” as very typical of DC-IC on the right side of the chart, for instance, one couldn’t help but wonder what on earth the first two terms had to do with the third. Then, we realized that we could push the analysis a step further: eliminate IC and IC-IC, for instance, and use principal component analysis to differentiate, not only between the various *sentence* types, but between their *clauses* as well. If semantic differences emerged *among* sentences, perhaps they would also emerge *within* them: between the dependent and independent clause of DC-IC, for instance; or between the independent clause of DC-IC, and that of IC-DC. And indeed, as [Figure 4.5](#) shows, a semantic separation occurred at this lower scale as well: the odd trio of “home”,

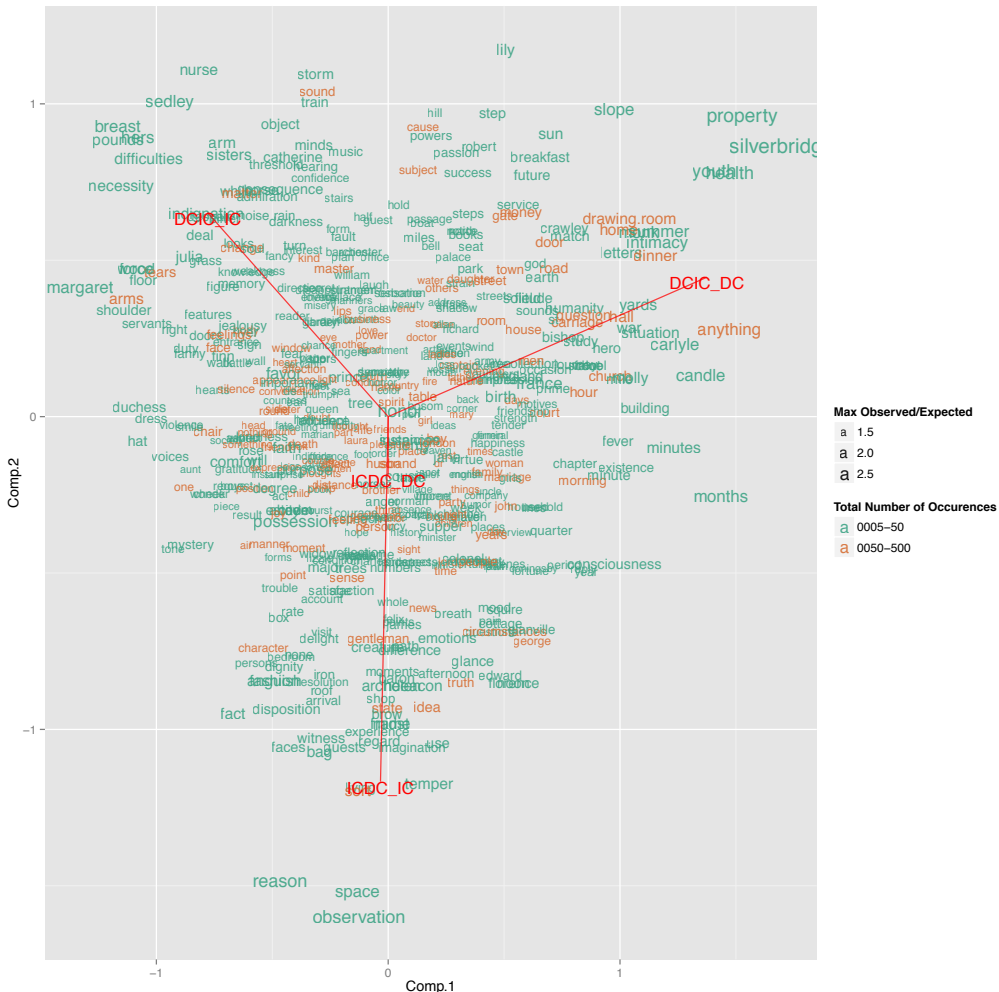


Figure 4.5: Two-clause sentences and their semantic spaces

“door”, and “change”, for instance—whose coexistence near the DC-IC vector of Figure 4.3 had so puzzled us—disaggregated into two very different semantic fields: “home” and “door” [plus “drawing-room”, “hall”, “church”, “gate”, “carriage”, “road” and other spatial terms] turned out to be typical of the dependent clause in the upper right quadrant of the chart, whereas “change” [plus “matter”, “feelings”, “indignation”, “despair”, “admiration”, “tears”] all clustered around the independent one, in the upper left quadrant. The more one looked, the clearer the semantic distance between the dependent and independent clause of DC-IC became.⁹ A correlation between grammar and semantics was thus beginning to emerge. Not a *necessary* correlation: rather a “line of least resistance”, as Jakobson had put it, in linking metaphors to poetry, and metonymies to prose; a “preference”, more

⁹ IC-DC sentences behaved somewhat differently: the independent clause possessed its own specific semantic pole (“idea”, “reason”, “observation”, “imagination”, “hate”), but the vector of the dependent clause was far less specified than the other three, as shown by its greater proximity to the center of the diagram. Given that these dependent clauses are often relative ones, which must be free to move in multiple semantic directions, their not being committed to any specific semantic domain seemed intuitively appropriate.

than anything else, that attracted space to the dependent clause of DC-IC, and strong emotions to the independent one. It was to this strange elective affinity emerging from the most “narrative” of our sentence types that we now turned our attention.

5. “When the procession came to the grave” [May 2012]

5.a

When the procession **came to the grave** the music ceased. [Radcliffe]

As they landed, a low growl of thunder was heard at a distance. [Scott]

As she **came out of church**, she was joined by Mr. Bellingham. [Gaskell]

As **I passed the steps of the portico**, I encountered, at the corner, a woman’s face. [Dickens]

When they **were in the streets** Esther hardly spoke. [Eliot]

Five different DC-IC sentences, from a large random sample that included very different authors, and always the same pattern: a spatial movement (bolded) occurs, *and then something else happens*: thunder growls, Mr. Bellingham joins the heroine, a woman’s face appears. If DC-IC sentences are in charge of narrative expansion, as we wrote a few pages back, these examples add the further specification that *a spatial movement in the dependent clause is often the springboard of narrativity*: first Ruth has to leave the church, and only then does the man who will ultimately seduce her approach. In the nineteenth century the *milieu* enters European narrative, of course, and space becomes a more tangible presence; but as the example of Auerbach’s *Mimesis* makes clear, the obvious place for space and the milieu are novelistic *descriptions*. And instead, we had found them at the very source—at the *microscopic* source, one could almost say—of narrative developments. It was strange.

It was strange. And when we turned to the *independent* clause of DC-IC sentences, a comparable surprise was awaiting us. As the DC-IC is the sentence of narrative intensification, and perhaps even acceleration, we expected its main clause—which is where intensity increases—to resemble the Radcliffe sentence quoted in 3.b (“While she looked on him, *his features changed and seemed convulsed in the agonies of death*”), or this one from Dickens: “As I watched him in silence, he put his hand into the corner at his side, *and took up a gun*”. And instead, this is what we found:

5.b

When the ceremony was over he blessed and embraced them all with tears of fatherly affection. [Radcliffe]

As he recovered from a sort of half swoon, he cast his eyes eagerly around. [Scott]

While he listened, she ended her grateful prayers. [Gaskell]

When Miss Dartle spoke again, it was through her set teeth, and with a stamp upon the ground. [Dickens]

When Esther looked at him she relented, and felt ashamed of her gratuitous impatience. [Eliot]

In case after case, the semantic center of gravity of the independent clause had much more to do with *emotions* (sorrow, gratitude, shame, anger...) than with guns or throes of death. When a narrative intensification occurred, in other words, feelings mattered much more than actions or events—or perhaps, more precisely, emotional intensity *was* the event. It was a second surprise. And a third quickly followed, when we shifted our attention from the two clauses taken separately, to their combination. Since the semantic centers of gravity of the two clauses were so completely different—spatial movement in the dependent clause, and the expression of emotions in the independent one—it made sense that, in general, one of them should occur while the other did not. But there were also quite a few cases in which both semantic clusters were simultaneously activated:

5.c

When Peter **perceived the village**, he burst into a shout of joy. [Radcliffe]

When he **came up to Butler again**, he found him with his eyes fixed on the entrance of the Tolbooth, and apparently in deep thought. [Scott]

When she had got **behind the curtain**, she jumped on her father's neck, and burst into tears. [Disraeli]

When the day came round **for my return to the scene of the deed of violence**, my terrors reached their height. [Dickens]

When she **had once got to the seat** she broke out with suppressed passion of grief. [Gaskell]

When she **reached home** she found Mrs Pettifer there, anxious for her return. [Eliot]

We read these sentences with a mix of perplexity and disappointment: they were so—clumsy. Perhaps, inevitably so: space and emotions—which express, respectively, the power of the “milieu”, and the melodramatic undercurrent of the age—are such heterogeneous entities that combining them in the same short sentence may be simply impossible. And yet, every now and then, something seemed to happen:

5.d

When Deronda **met Gwendolen and Grandcourt on the staircase**, his mind was seriously preoccupied. [Eliot]

But when he **came in**, she started up. [Gaskell]

Yet when he arrived **at Stone Court** he could not see the change in Raffles without a shock. [Eliot]

When **their hands fell again**, their eyes were bright with tears. [Eliot]

These are much more evocative sentences. Instead of being activated in a mechanically uninspired way, the relationship between space and emotions becomes sharp and dynamic: the “realism” of setting, and the “melodrama” of feeling inter-animating each other with an almost Balzacian energy. It was an interesting find, this meeting-place between the two main axes of the nineteenth-century imagination. But, once again, it wasn't really what we had been looking for. The strength and elegance of these sentences seemed to exceed the semantic peculiarities we had meant to study. Would we have to change direction one more time?

6. Verbs and Genres [May 2012]

Slightly. Now that we had quantified so many sentence-level features—number and types of clauses, sentence length, logical relations—we decided to make an explicit connection with the research conducted two years earlier in “Quantitative Formalism”, where we had shown that the usage of the most frequent words (MFW) in English (like “the” and “of”) can accurately distinguish genres. Could sentence-level choices also distinguish genres? In other words, do genres have sentence styles? To answer this question, we tested which of our quantified sentence features (such as length, clause use and number, verb tense, and mood) could be used to meaningfully separate texts by genre. Among these features, it turned out that verb tense and mood were the most successful at creating generic distinctions. [Figures 6.1](#) and [6.2](#) show a couple of PCA charts from this phase of our research, involving the *Bildungsroman*, the Gothic, and the Jacobin novel.

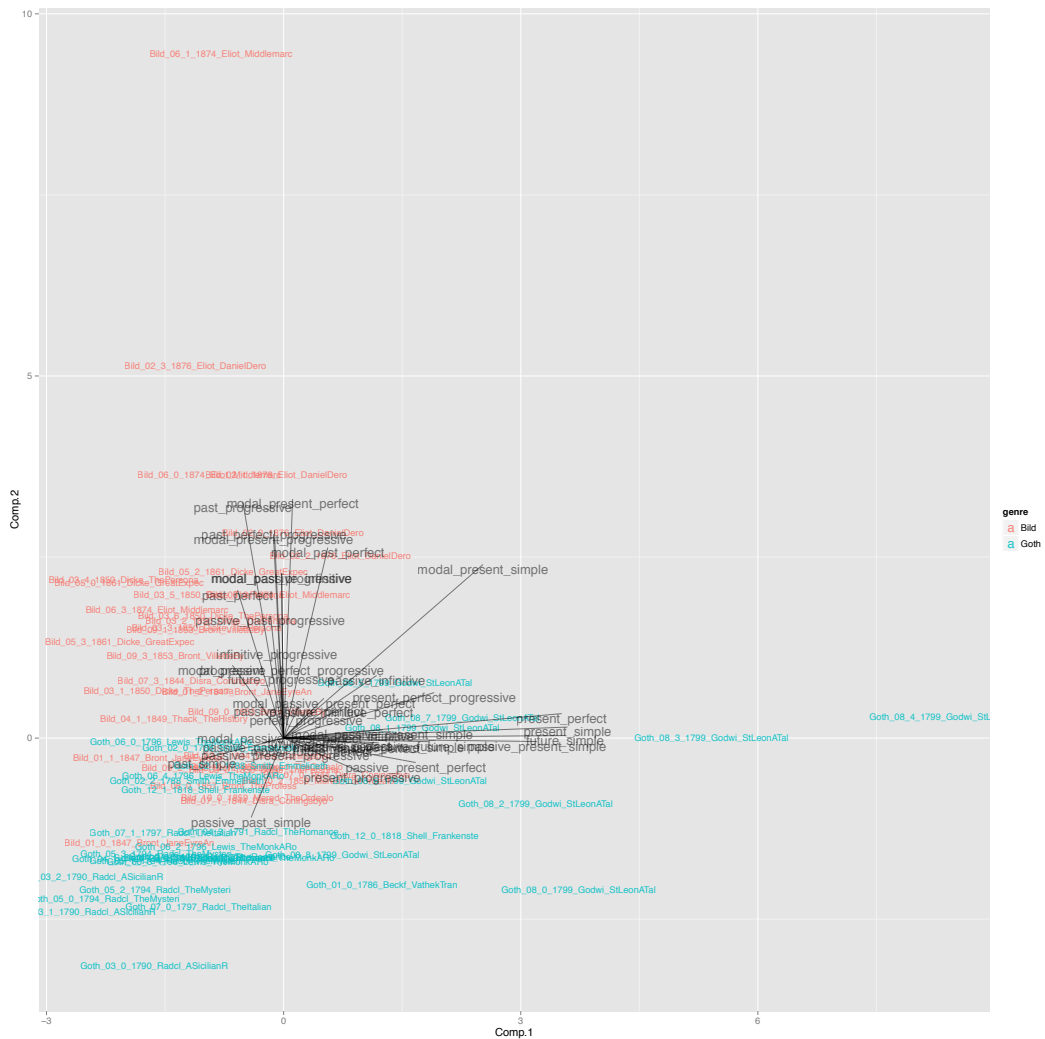


Figure 6.1: Verb forms as distinctive traits: Gothic novels and *Bildungsroman*

Figure 6.1 shows Gothic novels, in blue, clustering in the lower left quadrant, and the *Bildungsroman*, red, in the quadrant above; 6.2 shows Jacobin novels, in blue, spread along the horizontal axis left to right, and the *Bildungsroman* in the same location as 6.1. The units in the chart (Goth_03_0_1790_Radcl_AsicilianR in the bottom left corner of 6.1, or Bild_06_1_1874_Eliot_Middlemarc in the upper left one) are sections of the novels in our database, each of them containing 200 narrative sentences. The separation between the two genres is equally clear in both images, but the verb forms responsible for it are different: in the case of Gothic they are the perfect and the passive past simple, whereas for the Jacobin novel the key tenses are the present and the future. (For the *Bildungsroman*, on the other hand, modals and progressives remain constant as the key traits in both charts).

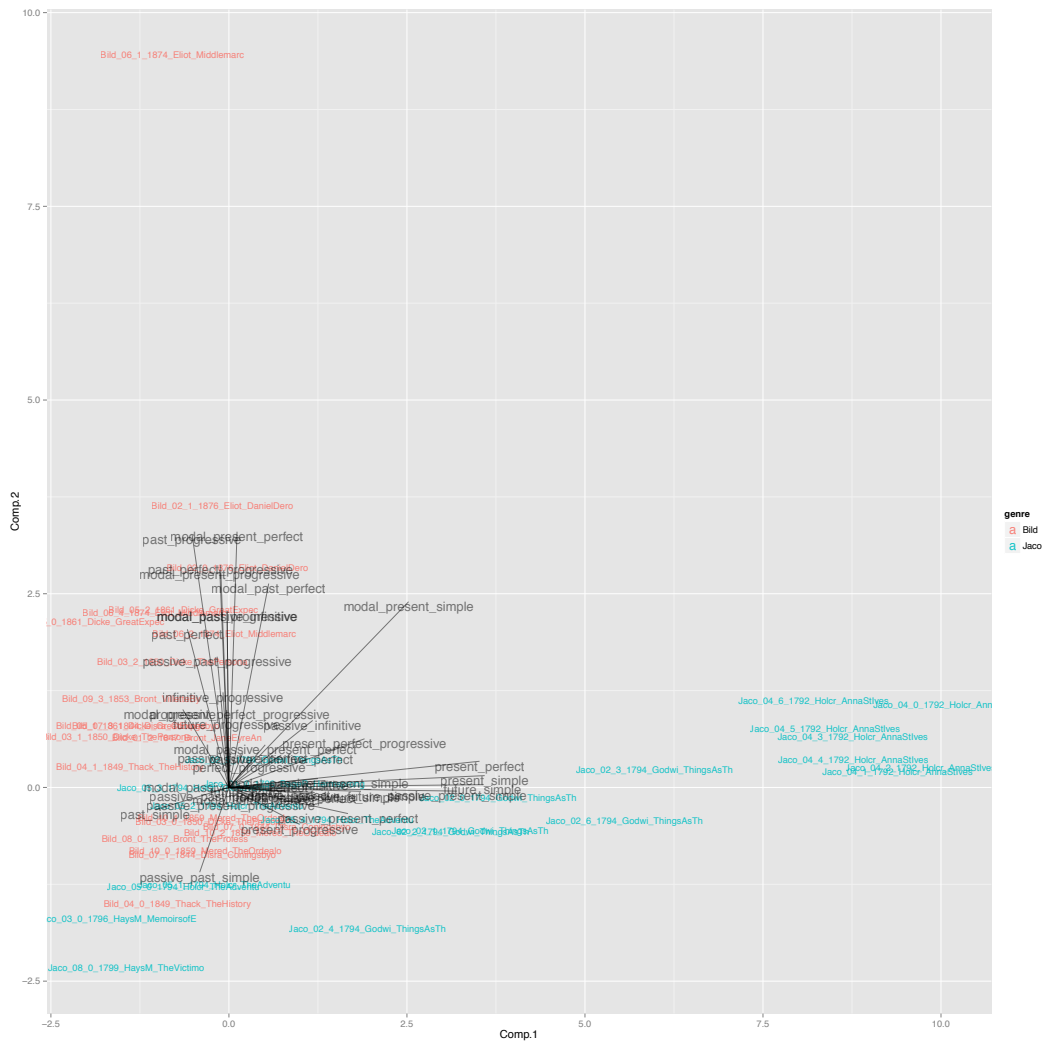


Figure 6.2: Verb forms as distinctive traits: Jacobin novels and *Bildungsromane*

Initially, we were taken aback by the fact that the past simple, or preterite, which a long theoretical tradition—from Benveniste to Barthes and Weinrich—had described as the fundamental tense of narration, played a negligible role in many of our charts. Placed slightly to the left of the point of origin of the vectors, along an axis where variance is seldom very strong, this tense contributes almost nothing to the separation between *Bildungsromane* and Gothic novels, which tends to be at its most dramatic along the vertical axis. On second thought, however, this lack of distinctiveness made sense: *precisely* because the simple past is the fundamental tense of narration, all novels use it quite often, and the increase in general frequency makes a strong variation from genre to genre unlikely. Perfect for telling apart novels from essays or scientific texts, the past simple is thus often useless in separating one novelistic genre from another. (Notice, however, that it *does* play a significant role in the case of Jacobin novels, with their strong orientation towards dialogue and away from narration).

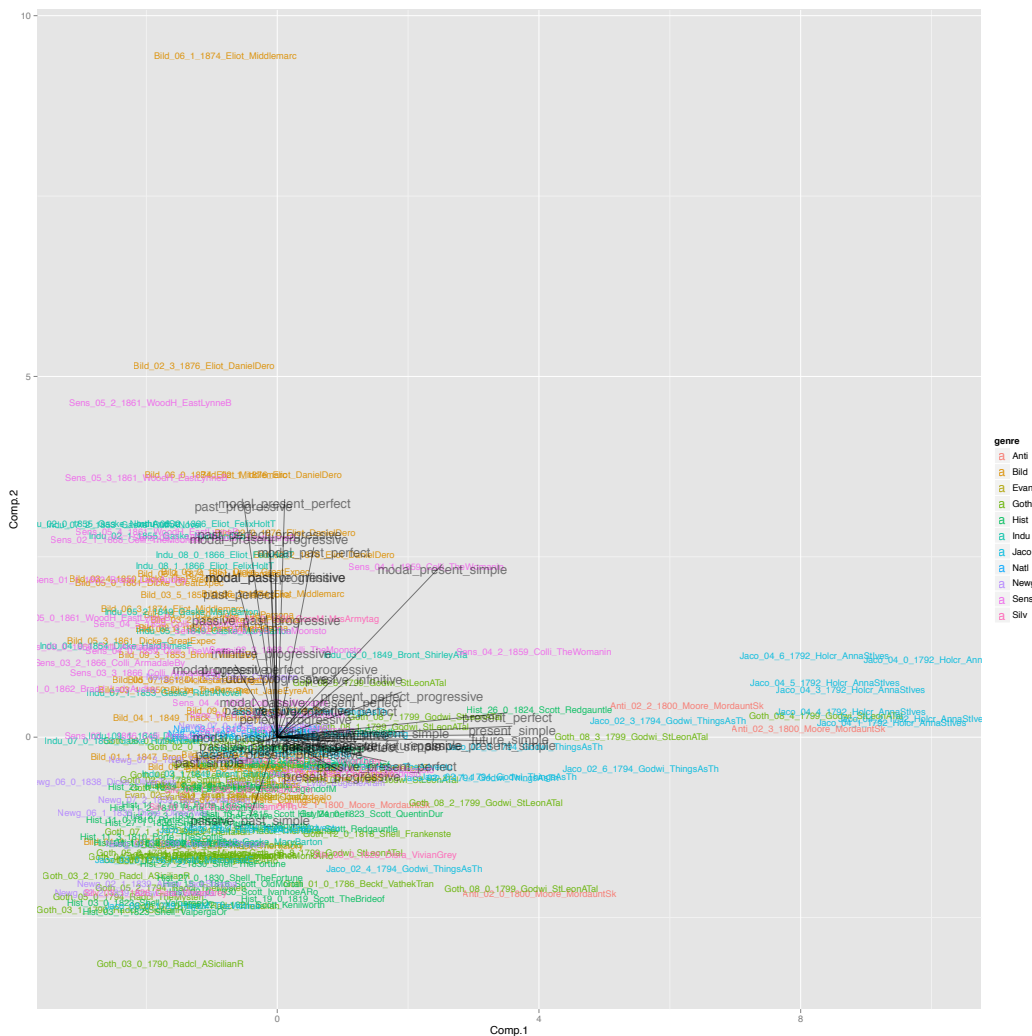


Figure 6.3: Verb forms as distinctive traits among eleven genres.

As chart followed chart, the results began to resemble those of “Quantitative Formalism”: in most cases the separation was good, in others less so—but none of the charts was truly surprising. Except, that is, for one detail that kept sticking out in chart after chart (including a mega-diagram in which we plotted eleven genres at once, mostly out of curiosity to see what would happen: Figure 6.3): segment 1 of *Middlemarch* was a total, almost ridiculous outlier. So we took the 200 sentences contained in that segment, and read them carefully.¹⁰ The PCA charts had already told us that we would find a large number of modals

10 Very carefully, in fact. Were we therefore doing a “close reading” of *Middlemarch*? Almost certainly not, for the simple reason that we were not reading *Middlemarch*, but a series which—as such—did not exist in the text, but was entirely an artifact of our methodology: an “artificial” object that “no one had ever seen and no one *could* ever see”, to quote Krzysztof Pomian’s *L’Ordre du Temps*. Although we try to be as attentive as any close reader to the details and formal properties of our sentences, the difference in the objects of analysis—a text, *versus* an artificial series—makes the use of the same term for the two practices quite misleading: “quantitative formalism” remains a much better description of our methodology.

and of progressive forms.¹¹ Now the question was: would this computer-generated series modify our understanding of the style of *Middlemarch*, or of that of the *Bildungsroman* as a genre? And if not, what was the point of the whole enterprise?

Let's begin with some instances of the progressive:

6.a

Mary was in her usual corner, **laughing** over Mrs Piozzi's recollections of Johnson, and looked up with the fun still in her face. It gradually faded as she saw Fred approach her without **speaking**, and stand before her with his elbow on the mantelpiece, **looking** ill. [...] She looked straight before her and took no notice of Fred, all the consequences at home **becoming** present to her. [...] Fred followed her with his eyes, **hoping** that they would meet hers, and in that way find access for his **imploring** penitence. [...] And when, **looking** up, her eyes met his dull **despairing** glance, her pity for him surmounted her anger and all her other anxieties.¹²

6.b

Lydgate, naturally, never thought of **staying** long with her, yet it seemed that the brief impersonal conversations they had together were **creating** that peculiar intimacy which consists in shyness. [...] They were obliged to look at each other in **speaking**, and somehow the **looking** could not be carried through as the matter of course which it really was.

6.c

And by a sad contradiction Dorothea's ideas and resolves seemed like **melting** ice **floating** and lost in the warm flood of which they had been but another form.

The sentences in 6.a come from the passage when Fred Vincy is about to confess to Mary Garth that he has lost a lot of money, thus creating serious financial problems for her family; 6.b, from the moment when Rosamond and Lydgate become aware that they may be falling in love with each other; 6.c, from the page when Dorothea's certainties about Casa-

11 As our examples will make clear, the forms in -ing are not always progressives, but often gerunds or present participles. Though all these forms imply an event in progress, thus justifying a single, synthetic analysis, speaking simply of "progressives" is a misnomer, which we have had recourse to for lack of a more descriptive general category.

12 The beginning of chapter 25 of *Middlemarch*, where these sentences appear, contains an extremely large number of forms in -ing:

"Fred Vincy wanted to arrive at Stone Court when Mary could not expect him, and when his uncle was not downstairs in that case she might be **sitting** alone in the wainscoted parlor. He left his horse in the yard to avoid **making** a noise on the gravel in front, and entered the parlor without other notice than the noise of the door-handle. Mary was in her usual corner, **laughing** over Mrs. Piozzi's recollections of Johnson, and looked up with the fun still in her face. It gradually faded as she saw Fred approach her without **speaking**, and stand before her with his elbow on the mantel-piece, **looking** ill. She too was silent, only **raising** her eyes to him inquiringly.

'Mary,' he began, 'I am a good-for-nothing blackguard.'

'I should think one of those epithets would do at a time,' said Mary, **trying** to smile, but **feeling** alarmed.

[...]

'Oh, poor mother, poor father!' said Mary, her eyes **filling** with tears, and a little sob rising which she tried to repress. She looked straight before her and took no notice of Fred, all the consequences at home **becoming** present to her. He too remained silent for some moments, **feeling** more miserable than ever."

upon begin to crumble. Very different situations, but with one trait in common: something important is about to happen—but hasn't yet fully crystallized. That's what the progressive (and the gerund) are for: presenting events as in progress, overlapping with others, not yet locked into a linear narrative: processes, more than results. Perfect, for the novel of youth: a season of life whose point lies in developing, changing, *becoming*.

And now the other verb form over-represented in *Middlemarch*—modals:

6.d

If a man could not love and be wise, surely he could flirt and be wise at the same time? [...] Now Lydgate might have called at the warehouse, or might have written a message on a leaf of his pocket-book and left it at the door. [...] A man may, from various motives, decline to give his company, but perhaps not even a sage would be gratified that nobody missed him.

6.e

In this solemnly-pledged union of her life, duty would present itself in some new form of inspiration and give a new meaning to wifely love. [...] She felt a new companionship with it, as if it had an ear for her and could see how she was looking at it. [...] She felt as if all her morning's gloom would vanish if she could see her husband glad because of her presence.

Here we are no longer in a fluid world of processes and transformations, but in one of uncertainty, politeness, and subdued emotions. In 6.d, for instance, the modals express Lydgate's erotic desire for Rosamond under the guise of impersonal and slightly ironical maxims: in part, a sign that Lydgate is not taking his own eros too seriously; in part, that he's ready to follow the *Middlemarch* notions of sexual decorum. By contrast, Dorothea's thoughts about her marriage in 6.e offer a version of youthful desire in which hopes too large to be expressed in the indicative ("duty would present itself...") morph sorrowfully into a "morning gloom", and are reborn only as a humble, hypothetical consolation ("would vanish if she could see"). In both cases, however, modals operate in a similar fashion: they take the open potentiality of the protagonists—youth, becoming—and overdetermine it with hermeneutic hesitation or the codes of social conformity.

Did the study of our series modify our understanding of *Middlemarch*, then, or of the style of the *Bildungsroman*? The first impulse was to answer both questions in the negative. That processes would be more important than punctual events, and possibility would mix uneasily with conventions—these were well-known features of the *Bildungsroman*, and it makes sense to find them visible in the novel's very diction. But a comparison with the analogous findings in "Quantitative Formalism" threw a different light on the matter. The spatial prepositions in the Gothic, we had written then, were clearly "consequences of higher-order choices": "effects of the chosen narrative structure"—of the desire to have a story where "every room may be full of surprises" ("Quantitative Formalism", 24). No one would ever say that spatial prepositions *make* a Gothic novel. But can progressives and modals make a *Bildungsroman*? Maybe, maybe not, but the question is a real one: these sentence-level choices don't just *descend* from the larger imperatives of the genre; they can plausibly play a causal role in creating its overall atmosphere, shaping the linguistic sensibility that makes readers intuitively grasp the "sense" of the form as a whole. Progres-

sives were “perfect” for the *Bildungsroman*, we wrote a few paragraphs earlier; perfect, yes, but not obvious: in theory, youth could have been defined just as well by a strong use of the future, for instance, rather than by the progressive. Once a writer starts using the latter, however, her decision has higher-order consequences: it emphasizes the (present) instability of youth over its (future) aims, clearly enough; it shifts the narrative center of gravity from the novel’s ending (which would be emphasized by a frequent recourse to the future) to its “middle”, which is where transformations occur. The scale of the sentence has a much more *constructive* role than anything we had encountered in “Quantitative Formalism”.

And then, there was something else. Modals and progressives, though they both distinguish the *Bildungsroman* from other genres, do so in very different ways. Progressives represent processes that are slow and perhaps inconclusive—but are definitely part of the reality of the plot: in 6.a, Fred is perplexingly silent, the cause for his “looking ill” is unknown, and the consequences of his act are incalculable: but there is no question that all these indeterminate events are actually *happening*. Modals, on the other hand, represent what is being *merely imagined* by the characters, and often jealously protected from public scrutiny. The two verb forms embody that great polarity of “world” and “soul” that is essential to the *Bildungsroman*: not just different, but *antithetical* dimensions. And yet, there are moments when the two suddenly converge:

6.f

His obligations to Mr Casaubon were not known to his hearer, but Will himself was **thinking** of them, and **wishing** that he could discharge them all by a cheque. [...] The allusion to Mr Casaubon would have spoiled all if anything at that moment could have spoiled the **subduing** power, the sweet dignity, of her noble unsuspecting inexperience. [...] If he never said a cutting word about Mr Casaubon again and left off **receiving** favors from him, it would clearly be permissible to hate him the more.

6.g

Moreover, Lydgate did not like the consciousness that in **voting** for Tyke he should be voting on the side obviously convenient for himself. [...] Other people would say so, and would allege that he was **currying** favor with Bulstrode for the sake of **making** himself important and **getting** on in the world.

The sentences in 6.f refer to Will Ladislaw’s attempt to balance his new desire for Dorothea with his old obligations towards her husband; those in 6.g, to the vote for the hospital chaplaincy, when Lydgate feels torn between the assertion of his own autonomy, and the incipient realization of the force of circumstances. In both episodes, the friction between reality, desires, and social norms is particularly harsh, and further strengthened by Eliot’s virtuoso use of free indirect style, which makes it hard to separate the character’s voice, social *doxa*, and the narrator’s judgment.¹³ Insofar as the *Bildungsroman* presents a grow-

13 The Lydgate passage goes on endlessly oscillating between one register and the other into the famous formulation: “For the first time Lydgate **was feeling** the **hampering** threadlike pressure of small social conditions, and their **frustrating** complexity. At the end of his inward debate, when he set out for the hospital, his hope was really in the chance that discussion might somehow give a new aspect to the question, and make the scale dip so as to exclude the necessity for voting. I think he trusted a little also to the energy which is begotten by circumstances—some feel-

ing entanglement—a “web”, in Eliot’s famous trope—of factual processes, subjective hopes, and symbolic norms, these sentences offered a true distillation of *Middlemarch* as a whole. The novel in one sentence, one was tempted to say.

7. Style at the Scale of the Sentence [Winter 2012—Spring 2013]

It was time for some final reflections. We had begun by imagining a study of style; then, our initial findings had made us switch our focus to narrative, and later to semantics; finally, the space-emotions continuum of DC-IC, and the orchestration of progressives and modals in *Middlemarch*—different as the two cases were—had created yet another scenario: elements that could exist perfectly well independently of each other (the delimitation of space, and the expression of emotions; the slow processes unfolding through progressives, and the hypothetical worlds conjured up by modals) showed a tendency to become amalgamated into powerful composite sentences. And then we realized that the scenario wasn’t new after all, since the sentence that had triggered our entire research—“Miss Brooke had that kind of beauty which seems to be thrown into relief by poor dress” (in the terms of this pamphlet, an IC-DC “defining” sentence)—was itself a product of the same mechanism: the narrative statement (Miss Brooke had a peculiar kind of beauty) and the essayistic specification (there is a kind of beauty which is emphasized by poor dress), could have perfectly well existed side by side, without interacting; once linked, however, their convergence into a single short statement made the opening of *Middlemarch* impossible to forget.

A completely unrelated project (Moretti’s book on *The Bourgeois*) provided one more instance of the same phenomenon. This time, the process came in three stages. While working on *Robinson Crusoe*, Moretti had noticed an unusual frequency of final clauses (IC-NFC) in Defoe’s novel, interpreting them as the stamp of “instrumental reason” over Robinson’s activity (“I did this, *in order to* do that”). Later, he noticed an even higher frequency of the mirror configuration (NFC-IC), in which the grammatical “aspect” of the past gerund suggested that mastery over the flow of time (“*having done this*, I then did that”) which is typical of Defoe’s novel. Finally, he had found several instances in which past gerund, main clause, and final clause (NFC-IC-NFC: “and *having stowed* my boat very safe, I *went* on shore to look about me”) were so tightly interwoven that the sentence’s very grammar, in its uninterrupted movement from past to present to future, seemed to embody that “forever renewed” activity that Max Weber had singled out as the psychological basis of capitalist accumulation. If there was a style of bourgeois laboriousness, Moretti had concluded, this was certainly it.¹⁴

If there was a style ... After having been abandoned, the concept had returned to the center of our research. Defoe’s interconnected clause chains, the modals and progressives of *Middlemarch*, the essayistic relative clause attached to a narrative statement, the space ing **rushing** warmly and **making** resolve easy, while debate in cool blood had only made it more difficult. However it was, he did not distinctly say to himself on which side he would vote; and all the while he **was inwardly resenting** the subjection which had been forced upon him. It would have seemed beforehand like a ridiculous piece of bad logic that he, with his unmixed resolutions of independence and his select purposes, would find himself at the very outset in the grasp of petty alternatives, each of which was repugnant to him”.

14 Franco Moretti, *The Bourgeois. Between History and Literature*, Verso, London 2013, pp. 37-9, 51-8.

and emotions in DC-IC: in every instance, a specific style had “emerged”¹⁵ from a process of syntactico-semantic condensation that was both *unexpected* and *reiterated*.¹⁶ Style was this condensation; that was why the frame of the sentence—and of *two-clause* sentences in particular—had become so important for us: it was the smallest linguistic construct whose parts could freely coalesce into new, emergent complexes, *making the genesis of style empirically observable*. Two-clause sentences were the laboratories of literary style.

Style as a “condensation” of discrete elements within a sentence, then. We had encountered so many instances of this process, and so varied, that we had no doubts as to its existence. But—why had the condensation occurred in the first place? Why had modals and progressives, or gerunds and final clauses, become so deeply associated within a single sentence? The best answer seemed to be: because they could. The semantic fields of space and emotions, or the verb forms of *Middlemarch* and *Robinson Crusoe*, were present in hundreds of clauses, which interacted in a variety of ways in the course of the novel: their encounter was thus firmly in the realm of the possible—of the “*adjacent possible*”, as Steven Johnson has called it, following Stuart Kauffman: good solutions that come into being not as inventions *ex nihilo*, but as lucky discoveries of a fertile relationship between already-given, and often widely-circulating ideas.¹⁷

The “adjacent possible” was a great formula to capture both the nature and the emergence of style. As something that was merely *possible*, style did not have to be there for a given text to exist more or less in its current form: Eliot didn’t *need* the “Miss Brooke” type of sentence in the same sense she needed to resolve the various marriage plots of *Middlemarch*. At the same time, that sentence belonged, not just to the realm of the possible, but of the *adjacent possible*: given Eliot’s passion for both story-telling and essayistic reflection, it was definitely likely to occur. Neither inevitable, nor truly exceptional, style occupied a middle position between logical extremes, where borderline cases were necessarily frequent: as in the case of the “is this style, or not?” of our initial discussion, or, of the elusive border between success and failure in the space-emotions combinations at the end of section 5. This was why a slightly indeterminate definition such as “emerging from the condensation of independent elements” was so appropriate: there are concepts—like “blue”, “bald”, or “tadpole”—that signify *through* a certain amount of vagueness, rather than despite it,¹⁸ and style is probably one of them. The uncertain boundaries of terms such as “condensation”, or “elements”, allowed us to see the peculiarity of the phenom-

15 Retrospectively, the concept of “emergence” explained the initial disagreement about reductionism. For reductionism, what happens at the most elementary levels of organization is really *all* that happens, and higher levels are simply a magnification of the basic processes; for the concept of emergence, by contrast, larger structures acquire properties that were not present in their separate components, and that therefore cannot be explained on their basis. This said, there is a significant difference between our use of the concept of emergence, and the one current in the natural and social sciences: in the latter, emergence indicates a process in which the agents involved (ants; passers-by; competitors on the market etc.), though continuously interacting, exist *independently* of each other in ways that are unimaginable for clauses and sentences belonging to the same text.

16 That style needs both deviation and repetition had been clearly formulated by Gemma in her dissertation: “I define style as a deviance from patterns that becomes a recurrent pattern itself; style emerges at the moment that it is both divergent enough from some norm to be noticeable and frequent enough to constitute a pattern of its own.” Marissa Gemma, *Exceedingly Correct: Stylistic Polemics in Nineteenth-Century American Literature*, Stanford University 2012, p. 27.

17 Steven Johnson, *Where Good Ideas Come From. The Natural History of Innovation*, Riverhead, NY 2010, p. 31ff.

18 See Rosanna Keefe and Peter Smith, eds., *Vagueness. A Reader*, MIT Press, Cambridge 1966.

enon of style, and of its process of formation: trying to cleanse them of all fuzziness, and of all borderline cases, would not make our understanding more precise, but destroy its possibility altogether.

Neither inevitable nor exceptional, style appeared as an eminently *comparative* fact: something that was not *necessary* to accomplish a given aim (which would amount to a functional definition of style), but that allowed it to be accomplished *better* than would otherwise be the case. Subject to all sorts of unpredictable contingencies, style never really *had* to emerge. When it did so, however, then it became immediately typical and recognizable: it distinguished an author, a genre, or a literary movement, in the most direct and unambiguous way. Here, the decisive category seemed to remain that of the author: a fact that had already become evident in “Quantitative Formalism” (pp. 13-15), and that returned in this study, with the growing role played by Defoe and Eliot in the final sections. Here, however, the relationship between author and genre revealed something that we had completely missed in the earlier pamphlet: what Eliot did with her modals and progressives, or the mix of narration and comment, did not *contradict* the logic of the *Bildungsroman* (the author *versus* the genre, as we had written in “Quantitative Formalism”), but rather expressed its central point with particular cogency (the author as *the highest embodiment* of the genre). If every *Bildungsroman* told the story of a young person, and explicated it, on a distinct textual level, with the voice of a reflectively adult narrator, *Middlemarch* showed the spark that arose from the direct encounter of the two planes. In Eliot’s sentence, two separate generic “traits” had turned into a structure.

Style as a process of condensation that transcended what was strictly functional and necessary. The adjacent possible as the source of the condensation, and the author/genre dialectic as its historical horizon. And the sentence? Had anything happened at the scale of the sentence *that could not have happened at any other scale*? Compared to other units analyzed by stylistics, the sentence’s brevity made it the perfect vehicle of textual *concentration*: taking the central meaning of a text, and compressing it in such a way as to make it unforgettable. And it’s not just a matter of brevity. When the “wisdom” of Eliot’s comments is conveyed by a relative clause seamlessly blended to a narrative statement, her values seem to emerge “naturally” from the story that is unfolding, rather than being an external reflection on it. When Robinson’s “forever renewed” activity is expressed by Defoe’s NFC-IC-NFC—“*Having mastered* this difficulty, and *employed* a world of time about it, I *bestirred* myself to *see*, if possible, how to supply two wants”—bourgeois work ethic becomes inscribed in the novel’s very grammar, and hence enormously strengthened. The message becomes twice as effective, because it manifests itself, not just as a specific statement, but as a repeatable *linguistic practice*. Bourdieu:

Structured structures predisposed to function as structuring structures, that is, as principles of the generation and structuring of practices and representations which can be objectively ‘regulated’ and ‘regular’ without in any way being the product of obedience to rules, objectively adapted to their goals without presupposing a conscious aiming at ends or an express mastery of the operations necessary to attain them.¹⁹

“Structured structures” that come into being by the slow accretion of distinct yet compatible elements; and that, without “consciously aiming” at doing so, “regulate” the reader’s

19 Pierre Bourdieu, *Outline of a theory of practice*, 1972, Cambridge UP 2012, p. 72.

“practice and representation” of temporality, or of possibility, or of ethical behavior. This is what style at the scale of the sentence can do: spacious enough to include a whole structure, it is also small enough to be easily grasped and absorbed, thus indeed “regulating” expression, as Bourdieu puts it. Style as *habitus*, in other words: as something that could spill over from grammar and literature into psychic structures and social interactions.

This broader, “social” notion of style, we concluded, deserved its own study. As for this pamphlet, its central contribution was simpler: a notion of style as combination and condensation, derived from the study of how clauses combine. That style resulted from the combination of originally discrete elements, Gemma and Heuser observed during our final exchanges, also meant that these elements could be independently formalized, and that a computer program could gather and measure them. Our “definition” of style also entailed, in other words, *a method for looking for it*: it was the beginning of a possible “operationalizing” of the concept. It was precisely as a result of our successful (though partial) operationalizing that we had so often found ourselves in front of striking patterns—in particular, correlations between syntax and narrative, and syntax and semantics—which our critical categories were incapable of explaining. The instructions we had given our programs—i.e., finding which elements of the sentence vary according to syntactic choices, like beginning the sentence with a dependent clause—had worked only too well, placing us in front of solid, clear, and intuitively significant data, which we didn’t quite know what to do with. This seemed to be one of the most revolutionary aspects of digital humanities: one in which the “digital” challenged the “humanities” with an entirely new type of problem.

But it was just as relevant, Gemma and Heuser added, that our programs *could neither detect nor explain the combinations that led to the emergence of style*: able to identify the separate parts of the process, the significance of their interactions eluded them, because they rested on conjectural connections between syntactical choices and broader cultural phenomena, like the sudden convergence of novel and essay, space and emotions, or past, present, and future. Here, the “digital” clearly needed the “humanities” to make sense of its findings. The two sides of the enterprise revealed their profound complementarity, which was also mirrored in the composition of the present pamphlet: where sections 2-4 had been devoted to the quantification and correlation of discrete elements; and sections 5-6, to the growing awareness that the patterns we had found required interpretations on a different plane. Without the concepts of the second half of the paper, the results of the first would have remained blind; and without the empirical content of the first part, the categories of the second would have remained empty. Only from their encounter did critical knowledge arise.

An encounter between concepts and measurements, then. And, as our iterative research shows, this encounter is a feedback loop wherein concepts inform measurements, and further measurements bring into play further concepts. Though there was something rigid—as well as partial—about our initial mandate to look for style at the level of the sentence, and to tie its occurrence to quantifiable phenomena, the interaction of concept and measurement had triggered a dynamic process in the course of which we had found, not only a new definition of style, but a definition which sharply differentiated our work from the great stylistic tradition of Spitzer and Auerbach. In their classic works—much richer, let us be clear, than anything we have done—the different components of style tend to add up, or to reiterate with minor differences the same general point: but they don’t *in-*

teract—let alone acquire emergent properties as a result of the process. Though far from incompatible, the two approaches study style at two different scales: that of the sentence, with its intuitively recognizable effects, and that of the paragraph, or of the text as a whole, with its often near-invisible touches. Unifying these two scales of the phenomenon (and of the concept)—here, a whole new pamphlet could definitely begin. And this one could, finally, come to an end.