Clausal order: A corpus-based experiment

Göran Kjellmer †, University of Gothenburg

Abstract
One variable feature among many in modern English is that of clausal order. The relative order of main and subsidiary clauses seems largely random, which is why a closer look into a representative corpus could be motivated. The present paper attempts to show with the aid of corpus material that the randomness element is more limited than expected and that fairly consistent tendencies on this score can be demonstrated.

Introduction

Different types of language variability have attracted the attention of linguists, particularly in recent times. To establish the extent to which a linguistic feature may vary is an important aspect of such studies. One way, out of many, in which a language like English displays seemingly random variation is in the ordering of the clauses making up a sentence. This paper is about such variation.

Consider the following sentences:

(1) Because I’ve been here so long I can cope, but the job has become a monster that nobody coming after me should be expected to handle.

(2) “Surely, you’re too big for a rocking-horse!” his mother had remonstrated. “Well, you see, mother, till I can have a real horse, I like to have some sort of animal about,” had been his quaint answer.

(3) As though it were a litany, Rabbi Binder repeated her words.
As if she sensed my observation she turned her face our way.

The adverbial subordinate clauses underlined precede their matrix clauses, but it is obvious that they can also follow them: “I can cope because I’ve been here so long”, etc., and in many cases postposition may even seem more natural, at least out of context. In fact, it is generally possible to switch adverbial finite clauses between pre and post position in relation to their matrix clauses. “Unlike non-finite clauses, finite [adverbial] clauses appear relatively often in both initial and final positions.” (Biber et al. 1999:835).

That some reservation is called for on this point is shown by examples like the following, featuring the conjunction *for*:

(5) I feared for my safety for he was trying to tear my office apart.

(6) Tony Blair cannot afford to lose her, for he would then stand isolated.

(7) She broke off, for he appeared to be chuckling.

(8) Shun that man for he is a part of the vast cosmos.

Any attempt to put the underlined clauses in pre position would result in unidiomatic or even incomprehensible English:

(5a) *For he was trying to tear my office apart, I feared for my safety.

---

1 See below about the Corpus.
(6a) *For he would then stand isolated, Tony Blair cannot afford to lose her.

(7a) *For he appeared to be chuckling, she broke off.

(8a) *For he is a part of the vast cosmos, shun that man

If it is not the case that the two positions of adverbial clauses, pre and post, are equally possible or equally frequent in relation to their matrix clauses, it should be of some interest to find out if there are any tendencies in this field, if different types of clause show different preferences. If there are clear preferences in this field, this may tell us something about how the clauses are used to convey information and, perhaps, about other issues of communicative interest.

Adverbial clauses are a subtype of adverbials, and adverbial position has previously been studied at length by a number of scholars, such as Jacobson (1964), Greenbaum (1969) and Ford (1993). The subject receives full treatment in Quirk et al. (1985) and, particularly, in Biber et al. (1999). It is clear from studies such as those that adverbial clauses cannot be assumed out of hand to have the same positions as adverbs in the sentence. An investigation of the placement of adverbial clauses in an extensive modern corpus like Cobuild may be able to contribute to our knowledge of the modern language, which is why this study was undertaken. No fine-grained analysis will be attempted; it is hoped that the categories offered by the Corpus will suffice to supply the basis for conclusions. Each adverbial clause type will be treated as a unit, and subtypes like subjunct, adjunct and disjunct clauses (Quirk et al 1985: 1069-1070) will not be distinguished within each category. Only finite clauses will be considered.

Material

The Cobuild Corpus was used as a basis for the present investigation. It consists of 56 million words from British, American and Australian sources, written and spoken. (See below for a list of the subcorpora
The clause-types singled out for closer study were finite clauses introduced by the following 18 subordinators:

after
although
as
as if
as though
because
before
even if
even though
for
if (conditional)
since
though
till
unless
until
when
while

300 examples of each clause-type were randomly selected from the Corpus, which then resulted in a corpus of (18x300=) 5400 finite adverbial clauses.

The selection criteria applied result in certain restrictions. Thus, as only clauses introduced by subordinators were included, non-introduced adverbial clauses were not included, as e.g. in

(9) I would have handled the situation in exactly the same way had I been in charge.

Further, nonfinite or verbless subordinate clauses are not included, such as
(10) Having equipment serviced or replaced takes place only when absolutely essential.
Corpus: ukmags/03. Text: N0000000577.

(11) Miser’s lips always trembled as though about to say something, but they very seldom did.

However, finite subordinate clauses whose matrix clauses are nonfinite are included:

(12) What if Ted’s world had no fixed horizons?

(13) So much for socialism today if we have councillors who can afford to go to their work wearing designer suits.

The subject may be omitted in a finite clause:

(14) My checklist runs as follows: ...

(15) I gather he has left now, though still keeps turning up.

Prepositional homonyms of the subordinators (after, before, etc.) were of course excluded.

The 18 clause-types studied here will have to be seen as examples of different types, rather than as an exhaustive account, as the list of clause-types is not complete. (Some types are excluded because they are too infrequent.\(^2\))

\(^2\) In order that has 79 occurrences in Cobuild, on condition that 70, providing that 35, in the event that 41.
Three main positions of the adverbial clause are distinguished, “pre”, “mid” and “post”, the positions being all relative to the position of the matrix clause. In addition, a number of unclassifiable clauses were called “0”.

**Pre clauses** occur “before the subject or other obligatory elements of the clause” (Biber *et al.* [1999: 771]). They are, e.g.,

(16) Unless something is done to improve the situation, we could be facing chaos.  

(17) While there are many blacks playing the blues today, the most popular music is electric city blues.  

The subordinate clause in “pre” position may be preceded by an adverbial:

(18) Robertson had risen from the ranks. For this reason, though he and Haig agreed on most military issues, a wide gulf always divided them. Corpus: ukbooks/08. Text: B0000000551.

(19) First of all, if you can’t spot the issues, you can’t score points on an exam. Corpus: usephem/05. Text: E9000000232.

(20) One day when I arrived there, Homer was not in his usual chair on the porch.  

Note that a matrix clause may itself be subordinated:

(21) Souter recalled that when he was a student adviser at a Harvard dormitory, a freshman asked him to talk to the freshman’s girlfriend. Corpus: npr/07. Text: S2000900914.
Mid clauses are those that occur in “all positions between obligatory initial and final clausal elements” (Biber et al. [1999: 771]):

(22) Children’s rights, unless they are associated with abuse, get a bad press in this country.

(23) Anyway, the London Symphony Orchestra (for it was they) subsequently sacked the stage manager...

Mid clauses include subordinate clauses inside cleft sentences:

(24) It is this period, when contractions are closely spaced and strong, that women typically find the most painful.

(25) it wasn’t till he’d be about four or five that he started filling out

(26) what happens <ZF1> when <ZF0> when you draw that kind of an event is that you have to learn to use spatial relationships in a remarkable way. Corpus: ukspok/04. Text: S0000000294.

Post clauses occur after their matrix clauses:

(27) We will remember him, for we have lost a warrior.
Corpus: ukmags/03. Text: N0000000638.

(28) I tried to call her after I checked in at the Churchill, but there was no answer at their apartment.

As the last example shows, a post clause may be followed by another clause.

As post clauses are also classified those whose matrix clauses are uttered by a previous speaker.
(29) <F0X> Erm you know that
<M0X> Yes.
<F0X> there’s never been a sense that a waitress was any less capable than a waiter.
<M0X> Yes yes. **Though it’s always** the head waiter not the head waitress isn’t it?

(30) <M01> <ZF1> What <ZF0> what makes you think that they never worked in South Africa?
<M07> Because <ZF1> they <ZF0> they pardon **they’re no worse off** even now.

Biber et al. (1999: 833-834) discuss this phenomenon:

> One of the most striking, though quantitatively small, differences in adverbial clause use across registers concerns adverbial clauses added to other-speaker main clauses in conversation. The face-to-face nature of conversation makes it possible for one participant to add circumstances on to another participant’s utterance. Often this occurs with conditional clauses, where a second speaker adds a condition qualifying the truth value of the first speaker’s assertion […]

Post clauses are also those where the speaker interrupts himself between matrix and subordinate clause:

(31) I’ll talk to Marcus about it,” I said, feeling that I had done all that convention demanded, and getting to my feet. **Though** of course, as you know, it’s Colonel Weston who is the senior of the two churchwardens.” Corpus: ukbooks/08. Text: B000000018.

0 clauses, finally, are those where classification is impossible for various reasons (slips of the tongue, interruptions, quotations, changes of tack, fragmented sentences). A few examples are:

(32) That’s what ‘When We Two Parted’ is about.
Corpus: ukmags/03. Text: N000000812.
(33) <M01> Right. So <ZF1> you’re saying it’s not just a question of experts
<M03> Yeah well you can say <ZGY>
<M01> but who’s experts they were. Yeah
<M03> argue against their <ZGY>
<M0X> Yeah
<M03> if they wanted well
<M0X> No
<M03> <ZGY> sink it in the middle of the sea the erm Shell <ZGY>
<M0X> Yes
<M0X> Mm
<M01> Right

Results

Relevant findings will be presented in tables below. As the material was collected randomly from the whole Corpus where the guiding principle was that the same number of every clause-type should be included, this makes it possible to compare the occurrence of adverbial clauses in different subcorpora. The terms used for the subcorpora are these:

- npr = US National Public Radio broadcasts
- today = UK Today newspaper
- times = UK Times newspaper
- usbooks = US books; fiction & non-fiction
- oznews = Australian newspapers
- bbc = BBC World Service radio broadcasts
- usephem = US ephemera (leaflets, adverts, etc)
- ukmags = UK magazines
- sunnow = UK Sun newspaper
- ukspok = UK transcribed informal speech
- ukbooks = UK books; fiction & non-fiction
- ukephem = UK ephemera (leaflets, adverts, etc)

Table 1 presents the evidence.
Table 1. Occurrence of adverbial clauses in different subcorpora

<table>
<thead>
<tr>
<th></th>
<th>Total adverbial clauses</th>
<th>Words in subcorpus</th>
<th>Total adv. clauses normalised</th>
</tr>
</thead>
<tbody>
<tr>
<td>ukephem</td>
<td>137</td>
<td>3124354</td>
<td>4.4</td>
</tr>
<tr>
<td>usephem</td>
<td>65</td>
<td>1224710</td>
<td>5.3</td>
</tr>
<tr>
<td>oznews</td>
<td>412</td>
<td>5337528</td>
<td>7.7</td>
</tr>
<tr>
<td>ukspok (1)</td>
<td>264</td>
<td>- 9272579</td>
<td>- 8.2</td>
</tr>
<tr>
<td>ukspok (2)</td>
<td>498</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ukmags</td>
<td>408</td>
<td>4901990</td>
<td>8.3</td>
</tr>
<tr>
<td>bbc</td>
<td>219</td>
<td>2609869</td>
<td>8.4</td>
</tr>
<tr>
<td>npr</td>
<td>275</td>
<td>3129222</td>
<td>8.8</td>
</tr>
<tr>
<td>times</td>
<td>540</td>
<td>5763761</td>
<td>9.4</td>
</tr>
<tr>
<td>today</td>
<td>511</td>
<td>5248302</td>
<td>9.7</td>
</tr>
<tr>
<td>sunnow</td>
<td>578</td>
<td>5824476</td>
<td>9.9</td>
</tr>
<tr>
<td>ukbooks</td>
<td>726</td>
<td>5354262</td>
<td>13.6</td>
</tr>
<tr>
<td>usbooks</td>
<td>767</td>
<td>5626436</td>
<td>13.6</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>5400</strong></td>
<td><strong>54293135</strong></td>
<td><strong>9.9</strong></td>
</tr>
</tbody>
</table>

The column called “Total adverbial clauses” gives the raw figures for the number of times each subcorpus contains adverbial clauses in the sample. As the subcorpora are represented by unequal quantities of text, the totals have been made comparable by dividing the clause total for each subcorpus with the number of words and multiplying the result with 100,000.

The result of the calculations is quite striking. Subcorpora differ markedly in their use of adverbial clauses. The two “literary” subcorpora, ukbooks and usbooks, are way ahead of the others in their use of adverbial clauses. One can only speculate about the reason why this should be so, but it does seem probable that literary texts have a greater need for modification, reservation, concession, explanation, condition. Things are rarely black and white in fiction (or in life, come to that). Adverbial clauses are three times as frequent in the literary texts as in official types of writing (ukephem, usephem), where it is in the nature of things that modification is less often called for, the official nature of the material imposing certain restrictions. It may also be noted that there is no difference in adverbial-clause frequency between British and American types of material: ukephem and usephem are very close, the
radio stations bbc and npr are also very close, and ukbooks and usbooks are as close as can be. Oznews (Australian newspapers), finally, favours a more straightforward, matter of fact type of writing than the British newspapers (times, today and sunnow); oznews uses few adverbial clauses and is close to the ephemera subcorpora.

Let us now look at the overall distribution of the 5400 clauses over the positional categories, presented in Table 2:

<table>
<thead>
<tr>
<th>Position</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre</td>
<td>1272</td>
<td>23.6</td>
</tr>
<tr>
<td>Mid</td>
<td>55</td>
<td>1.0</td>
</tr>
<tr>
<td>Post</td>
<td>4044</td>
<td>74.9</td>
</tr>
<tr>
<td>0</td>
<td>29</td>
<td>0.5</td>
</tr>
<tr>
<td>Total</td>
<td>5400</td>
<td>100.0</td>
</tr>
</tbody>
</table>

It appears from the table that post position is the most frequent one with adverbial clauses: three out of four follow their matrix clauses. This tallies with earlier observations. In Jacobson’s (1964) material, post position is less frequent, 55% (p.106), but still the most frequent option. According to Quirk et al. (1985: 1037), “[c]lauses that are constituents of phrases almost always occur at the end of the phrases.” Biber et al. (1999) present no overall figures but say (p. 833), “Whereas all types of non-finite clause are uniformly preferred in final position, different types of finite clause are distributed in different ways.” It could be noted, in addition, that mid position is infrequent, only 1 per cent.

Do subcorpora also differ in their general positioning of the adverbial clauses? Table 3 shows the distribution of pre and post positioned clauses over the subcorpora.

---

3 66 books by mid-20th century British authors, one half of which is fictional and one half nonfictional.
The table shows that, although post position is everywhere preferred, pre position is quite frequent in the American subcorpora npr and usephem, every third adverbial clause being pre posed, while pre position is about average in the equally American usbooks. The fact that sunnow, the Sun newspaper, has a very low pre percentage in relation to ukmags, times and today makes one suspect a Sun house style restricting the use of pre positioned adverbial clauses.

If the variation between subcorpora with regard to pre and post position is moderate, the differences between the clause types with regard to their position in relation to their matrix clauses is again striking. Table 4 shows the frequency in percentages with which the 18 clause types occur in pre position.
Table 4. Position of adverbial clauses (per cent)

<table>
<thead>
<tr>
<th>Clause</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>For</td>
<td>0</td>
</tr>
<tr>
<td>Till</td>
<td>1</td>
</tr>
<tr>
<td>As though</td>
<td>2</td>
</tr>
<tr>
<td>As if</td>
<td>3</td>
</tr>
<tr>
<td>Until</td>
<td>6</td>
</tr>
<tr>
<td>Because</td>
<td>7</td>
</tr>
<tr>
<td>Before</td>
<td>12</td>
</tr>
<tr>
<td>After</td>
<td>15</td>
</tr>
<tr>
<td>Unless</td>
<td>24</td>
</tr>
<tr>
<td>As</td>
<td>26</td>
</tr>
<tr>
<td>Even though</td>
<td>30</td>
</tr>
<tr>
<td>While</td>
<td>31</td>
</tr>
<tr>
<td>Since</td>
<td>33</td>
</tr>
<tr>
<td>When</td>
<td>33</td>
</tr>
<tr>
<td>Though</td>
<td>40</td>
</tr>
<tr>
<td>Even if</td>
<td>47</td>
</tr>
<tr>
<td>Although</td>
<td>56</td>
</tr>
<tr>
<td>If (cond.)</td>
<td>57</td>
</tr>
</tbody>
</table>

As Tables 1 and 3 suggest, most of the adverbial clause types prefer post position. Figure 1, based on Table 4, illustrates the fact that the 18 clause types can be seen to fall into three main groups with regard to relative position. There is some degree of semantic consistency in each group. The first group, 1-8, rarely, if ever, occurs in pre position and consequently normally occurs in post position, mid position being a rare occurrence. The group is dominated by time clauses (till, until, before, after),

---

4 “Adverbial clauses of time ... may be placed before or after the main clause. After the main clause is the more neutral position.” (Carter and McCarthy2006:561)

5 “Reason clauses may come before or after the main clause:” (Carter and McCarthy2006:561)
clauses (as though, as if) also belong here. The second group, 9-14, also prefers post position but not as markedly as the first group. It is characterised by a number of bi- or multifunctional clause types: as (manner, time, reason, comparison\(^6\)), while (time, concession/contrast\(^7\)), since (time, reason\(^8\)) and perhaps when (time, concession\(^9\)). The time element is prominent, a fact that we will come back to below. The last group, 15-18, is more undecided when it comes to relative position. In fact, two of them, although clauses and conditional if clauses, even show a certain preference for pre position. This partly agrees with Jacobson’s (1964: 95) finding that in his corpus (c. 40,000 words) conditional clauses differ from other adverbial clauses in being more frequent in pre than in post position. On the other hand, Carter and McCarthy (2006: 562) lump together if and unless clauses and state: “Conditional adverbial clauses may be placed before or after the main clause. After the main clause is the more neutral position.” Our last little group is made up of clauses of concession and condition. One might therefore have expected to find even though and unless clauses in this group rather than in the second one. However, while the position of even though clauses is still unclear, the obvious difference in position between unless and conditional if clauses, where thus unless clauses much more often occur in post position, may be explained by the tendency of unless, but not of if (not), to introduce an afterthought, something added to the main statement, as in She hasn’t got any hobbies – unless you call watching

\(^6\) “Conversation and academic prose most commonly use as to show manner relationships, while fiction and news more commonly use as to express time.” (Biber et al. 1999: 846)
\(^7\) (On concessive clauses:) “After the main clause is the more neutral position.” (Carter and McCarthy2006:562). “Almost all occurrences of while as a subordinator in conversation express time. In contrast, over 80% of the occurrences in academic prose marks concession/contrast.” (Biber et al. 1999: 849)
\(^8\) “Conversation and news more commonly use since for time. Academic prose and, less markedly, fiction more commonly use since for reason.” (Biber et al. 1999: 849)
\(^9\) Cambridge Advanced Learner’s Dictionary (2003): “when ALTHOUGH /wen/ conjunction despite the fact that: He says he hasn’t got any money when in fact he’s got thousands of dollars in his account.”
TV a hobby or Have a cup of tea – unless you’d prefer a cold drink (OALD). Here the unless clause is different from if clauses in being less well integrated in the superordinated structure.

Figure 1. Pre position of adverbial clauses (per cent)

Although the existence of multifunctional clause types is likely to blur what tendencies there are in the material, it seems possible to suggest that clauses of time rarely take pre position and that, at the other end of the spectrum, clauses of condition and concession taking pre and post position are about equally frequent.

The multifunctional clause types thus blur the picture. It is not unlikely that their different functions tend in different directions, and that the overall percentages therefore reflect a cancellation of the differences.
The common denominator of the multifunctional clause types in the middle category is the time function. It is possible, then, that the clauses in question are used differently with regard to relative position when they are time adjuncts from when they are not. To see if that is so a simple experiment was done. 50 occurrences each of *as, since* and *while* clauses were randomly extracted from the Corpus, and the functions of the finite clauses were analysed. The result is shown in Table 5:

<table>
<thead>
<tr>
<th></th>
<th>Time pre</th>
<th>Time post</th>
<th>Non-time pre</th>
<th>Non-time mid</th>
<th>Non-time post</th>
<th>Irrelevant</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>As</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>9</td>
<td>33</td>
<td>50</td>
</tr>
<tr>
<td>Since</td>
<td>6</td>
<td>9</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>28</td>
<td>50</td>
</tr>
<tr>
<td>While</td>
<td>2</td>
<td>11</td>
<td>6</td>
<td>0</td>
<td>12</td>
<td>19</td>
<td>50</td>
</tr>
<tr>
<td>Total</td>
<td>8</td>
<td>24</td>
<td>12</td>
<td>2</td>
<td>24</td>
<td>80</td>
<td>150</td>
</tr>
</tbody>
</table>

Most of those called “Irrelevant” in the table consist of nonfinite clauses, which had to be excluded for the sake of consistency with the main enquiry. The rest of the material shows that *as, since* and *while* clauses used as time clauses differ from *as, since* and *while* clauses in other functions. As time clauses they occur in pre position clearly less often, $8/(8+24) = 25\%$, than they do in other functions, $12/(12+2+24) = 32\%$. The second group of adverbial clause types can now be seen to represent several contrasting tendencies, one towards post position, and one towards a pre/post equilibrium.

*Influencing factors*

Biber *et al.* (1999: 835-838) present three factors of importance for deciding the relative position of adverbial clauses, namely (a) cohesion and information structuring; (b) framing subsequent discourse; and (c) structural considerations. The second factor is difficult to apply to whole categories of clauses – and in this study we are dealing with clause types in bulk -- but the first one seems intuitively capable of explaining some
of the differences between our clause types. The factor implies that subordinate clauses in pre position tend to contain given information, while the main clause presents new information, and, conversely, when the main clause contains given information, the adverbial clauses, with new information, tend to be in post position (835). It seems reasonable to think that time and reason clauses regularly contain new information without which the sentence would not make sense, and that they therefore come last in the clausal sequence. That does not happen every time, as is illustrated by example (1), but sufficiently often to form a pattern and to make an example like (1) stand out. That conditional and concessive clauses are used indifferently to present given and new information also seems very plausible. The third factor, too, is relevant in a more general way when the relative positions of matrix and subordinate adverbial clauses are concerned. Adverbial clauses normally contribute important information to the sentential proposition, and the principle of end-weight therefore stipulates that they should follow their matrix sentences. Cf. Quirk et al. (1985: 920):

When two coordinated units are placed in sequence, the second unit gains focal prominence from its position (cf 18.3ff.). This prominence in terms of information focus also attaches to the final element in a subordination relation, but in the latter case the positional highlighting is combined with a highlighting based on the formal inequality of subordination.

The reason why conditional clauses and concession clauses, particularly although clauses, marginally favour pre position is thus probably that they often contain given information whereas the new information is supplied by the matrix clause. There could, however, also be another reason. With conditional clauses, it often happens that the proposition of the main clause cannot be accepted or understood until the condition or constraint presented in the subordinate clause has been presented and accepted. The hearer/reader must take in the information in the subordinate clause in order to understand the information in the main clause:

(34) If you just knew the minds of the players going into the game at the kickoff, you wouldn’t be surprised, you wouldn’t be upset.
(35) If they leave, then they know full well that they’re not going to be paid.

(36) If the contract is accepted tomorrow as expected, all the strikers will be rehired.

Something similar is true of although-clauses, which are often necessary for the matrix clauses to be fully understood.

(37) Although 94 percent of the children in the survey were smacked and disliked the experience, there was a high degree of acceptance of this form of punishment.
Corpus: oznews/01. Text: N5000951004.

(38) But although Brocket was guilty, it is not a wife’s duty to betray. Rather the opposite.

(39) Although business volumes were up, confidence though now increasing showed a smaller rise.

In cases such as these the if and although clause prepares the reader/hearer for the message in the matrix clause and the sentential proposition is instantly accepted. If such necessary adverbial clauses should occur in post position, the message in the matrix clause would be hanging in the air, metaphorically speaking, waiting for the end of the sentence to be interpreted.

**Summary and conclusions**

Although adverbial clauses can generally precede or follow their matrix clauses, it was suspected that their distribution over pre and post position was not entirely random. 300 occurrences each of 18 adverbial clause
types were excerpted from the Cobuild Corpus and analysed with regard to position in relation to their matrix clauses.

The Corpus is made up of 12 subcorpora representing different styles and origins. It appeared that there were great differences between them in the occurrence of adverbial clauses, which were frequent in the “literary” subcorpora and infrequent in the more formal and official ones. When it comes to pre or post position (mid position is very infrequent), it was seen, first of all that, over all, three quarters of the adverbial clauses occurred after their matrices and that there was moderate variation in that respect between the subcorpora. On the other hand, the clause types differed greatly among themselves in positional tendencies, varying from no pre occurrences or very few (for and till clauses) to more than 50% such occurrences (although and conditional if clauses). Time and reason clauses occurred at one end of the spectrum, where post position was definitely preferred, and condition and concession clauses were found at the other end, where pre and post position were about equally common. Those tendencies were reinforced when a sample of as, since and while clauses, showing contradictory tendencies, were analysed separately.

With reference to Biber et al. (1999) it was suggested that information structuring (given and new information) and the principle of end weight influenced the relative positioning of the clauses. A special type of information structuring was the comprehensibility factor, i.e. the factor influencing which part of the sentence needs to be processed first in order for the proposition to be understood.

The general impression left by the survey is that the sequence of clauses is far less random than a superficial look would lead one to think, and also that the tools supplied by the Corpus can take you a long way.

References

Göran Kjellmer


CobuildDirect Corpus, an on-line service: telnet://titan.collins.co.uk


