# EXPLORING PARAGRAPHS FROM THE OUTSIDE

*Renata Pípalová*

**Abstract**

The paper strives to examine some conspicuous external parameters of graphic paragraphs in contemporary British English. It is based on a corpus composed of 2,070 paragraphs representing three register corpora and six subcorpora. The study surveys mainly register-specific features of paragraphing and discusses an interplay of factors affecting paragraph length and related phenomena.

## 0 Introduction

### 0.1 Graphic paragraphs as punctuation marks

Many authors include paragraphing among punctuation devices (e.g. Carey 1976, King 2000, Quirk et al. 1985). However, as King (2000: 16) notes, of all the marks of punctuation, the paragraph is the least precise and the most resistant to rules.

In fact, paragraphs may, but need not, be indented. In the non-indented cases, so-called American paragraphs, paragraphing is signalled by a skipped line. In the other group, the first line tends to be indented approximately five spaces if the paragraph is typewritten, and one inch if it is handwritten (e.g. Hodges et al. 1986, Troyka 1987). As a rule, the initial paragraph immediately following a title or subtitle is usually not indented at all (e.g. Macpherson 1997, Quirk et al. 1985). Indenting may be accompanied by other features, such as capitalization of the initial word, augmenting and/or decorating the very first letter, etc. Occasionally, to signal the position of paragraphs more unambiguously, authors may employ various techniques of numbering and lettering (see e.g. Bečka 1992). It should be noted that the end of paragraphs is not standardized.

### 0.2 Paragraph length

Indeed, the length of graphic paragraphs ranks among their most conspicuous features. Some authors associate it partly with eye-appeal (e.g. McArthur 1992). Carey (1976) stresses that pages divided into paragraphs at fairly frequent intervals are "more inviting to the eye than those run on with never a paragraph from top to bottom" (ibid.: 88). Wichman (2000: 30) goes further and contends that paragraph length is dictated mainly by aesthetic considerations.

When dealing with paragraph length, some authors observe the tendency towards uniformity. Dubois (1997) mentions two such studies, namely Lewis (1970) and Bond and Hayes (1984). The latter co-authors verified the tendency towards uniform paragraph length empirically. Their subjects enclosed on average 3.9 sentences in paragraphs. If the sentences were very long, their number per paragraph fell down to two.

Many other authors, however, stress that the length of paragraphs can vary a lot. As King (2000) observes, we sometimes set off a single word, but sometimes we see "leviathan examples which take up a page or more" (ibid.: 16).

Various studies discuss factors affecting paragraph length: Some mention the role of register (e.g. Brown & Yule 1983). According to Quirk et al. (1985), "short paragraphs provide convenient resting places during reading and aesthetically pleasing contrasts between print and white space. Both considerations account for the shorter paragraphs in the columns of newspapers and magazines, the more popular publications tending to have the shortest paragraphs" (ibid.: 1624). Others associate length with intentionality. For instance, Mathesius (1942 [1982]) writes that the length is determined by the author's intention and taste. According to McArthur (1992), longer paragraphs may be divided up for purposes of highlighting or emphasis. According to Daneš et al. (1987), the length of paragraphs depends on the text type, type of paragraph, mutual relations between paragraphs, author's intention and functional style. Boundaries between paragraphs often coincide with boundaries between direct and indirect speech, etc. The beginnings as well as ends of paragraphs are places especially exposed to constituting and maintaining textual connectedness.

Authors in the rhetorical tradition also generally admit that paragraphs vary in size (e.g. Heffernan & Lincoln (1982). Macpherson (1997) goes even as far as quantifying the ideal length of paragraphs: "There is great freedom as regards length, though anything between 50 and 150 words might be considered optimal" (ibid.: 9). Warriner et al. (1977) claim that "the paragraph may be long or short, depending on the idea it develops". Kolin (1986) urges writers to avoid excessively long paragraphs and to string together a series of short paragraphs. He believes that visual clues should be used for emphasis and organization and recommends that paragraphs should be visually attractive (suggesting a length of 8-10 lines). Stanford (1993) claims that most well-developed paragraphs should range between four to ten sentences.

Furthermore, many authors have studied the stylistic impact of paragraph length. For King (2000), "long paragraphs are tedious but short ones are jerky and can be equally hard to follow" (ibid.: 16).

Scrutinizing samples of Slovak fiction, Miko (1973) arrives at a very comprehensive formulation of the stylistic impact of paragraph length. In his view, long paragraphs allow the writer to dwell on a particular subject matter at length. The

resulting pace of the text is peaceful. Therefore, articulation into lengthy paragraphs is employed in the "stream of consciousness" technique. In addition, a low degree of text segmentation relaxes text organization and embodies the writer's resignation to interfere into the subject matter.

Conversely, Miko (ibid.) believes that shortening of paragraphs produces a kaleidoscope-like view of reality, as if through camera shots. Moreover, short paragraphs are associated with shorter prose formats. They may connote motion, excitement, tension of the situation, dynamicity or unrest. Paradoxically enough, however, sometimes extremely short paragraphs, especially descriptive ones charged with emotive content, may help create a rather static, and motionless atmosphere. Interestingly, a similar tendency is observed by Mistrík (1985).

## 1 Research corpus

The current paper strives to explore some conspicuous external parameters of graphic paragraphs in contemporary British English. It is based on a corpus comprising eighteen authentic sources, representing three register corpora and six subcorpora.

More specifically, the overall corpus consists of three corpora corresponding to distinct functional styles/broad stylistic varieties/registers[1], namely academic writing, journalism and fiction. Each register corpus is further divided into two subcorpora. The academic corpus includes a subcorpus from the humanities and one from the natural sciences; the journalistic corpus is made up of newspaper and magazine subcorpora, and the fiction corpus comprises children's as well as adult literature subcorpora. Naturally, by exploring several sources per (sub)corpus, the potential impact of the authors' idiolects was reduced.

Turning now to the quantitative aspect, all in all, the overall research corpus includes 2,070 paragraphs of contemporary British English. It is made up of three register corpora, each comprising 690 paragraphs. Every register corpus is in turn composed of two subcorpora, each featuring 345 paragraphs. Every register subcorpus includes three source subcorpora. Finally, the share of the sources is even, as each source embraces 115 continuous graphic paragraphs. The list of the corpus sources (all published within the last two decades), together with their abbreviations, is provided in the Primary Sources at the end of the paper.

### 1.1 Corpora and subcorpora in terms of length

The overall corpus of 2,070 paragraphs corresponds altogether to 7,452 utterances (hereinafter only U), i.e. 9,295 main clauses (hereinafter only C), see below. However, the distribution of the U/C among the register corpora and

subcorpora, among the source subcorpora, and, after all, as we shall strive to show below, also among the individual paragraphs, is far from even. In other words, although the (sub)corpora exhibit a fixed number of paragraphs, their absolute lengths vary. For example, though the register corpora consist of 345 paragraphs each, the academic corpus involves 3,225 U (i.e. 3,937 C), the journalism corpus consists of 2,010 U (i.e. 2,325 C), and the fiction corpus includes 2,217 U (i.e. 3,033 C).

### 1.2 Corpus in terms of samples

It should be pointed out that since the overall corpus comprises various text-types, some of the eighteen source-subcorpora feature only a single text sample (e.g. extract of a novel, academic monograph, etc.), whereas others are naturally composed of more text samples (e.g. newspaper articles, short stories, fairy-tales, legends). Source subcorpora comprising a single sample include A1, A2, A3, A4, A5, A6, F2, F4, F5, and F6. Conversely, multi-sample ones are J1, J2, J3, J4, J5, J6, F1, F3. This shows that the two groups are relatively balanced. An overview of the numbers of samples together with a more specific discussion of their impact is provided below.

### 1.3 Survey of samples enclosed in source subcorpora

***Academic corpus:***
**Natural Sciences subcorpus**
**A1** a single continuous sample (reference book), pp.17-32
**A2** a single continuous sample (monograph), pp.8-82
**A3** a single continuous sample (popularizing text), pp.1-29
**Humanities subcorpus**
**A4** a single continuous sample (monograph), pp.143-183
**A5** a single continuous sample (textbook), pp.10-33
**A6** a single continuous sample (monograph), pp.12-56

***Journalism corpus***

**Newspaper subcorpus**
**J1** six samples as follows (dates/titles of articles): 4/1/02 *Criminals Benefit from Euro Launch*; 5/1/02 *Dolly Hobbles back into the Limelight*; 25/11/03 *What is Killing our Children?;* 28/11/03 *Government Paves the Way for "New New" Universities;* 29/11/03 *Launch of the Annual Guardian Charity Appeal; High Times in Magic Mushroom Business – and It's Perfectly Legal;*

**J2** ten samples as follows (dates/titles of articles): 4/1/02 *Cause of Dyslexia Narrowed down to Single Chromosome; Pioneering Teachers to Get Laptops and Support Staff*; 5/1/02 *Dolly's Arthritis Raises Fear of Fast Ageing in Clones;* 28/11/03 *UK Obesity Epidemic not our Fault, Say Food Firms*; 29/11/03 *British Universities Attracting China's Best and Brightest; Birmingham Tempts Foreign Tourists; Rail Passengers Warned of Disruption on Busy Christmas Routes because of Track Closures; Cancers Rise despite Billions Spent on Research*; 30/11/03 *A Year of Weather Extremes. More of the Same to Come?; Revealed: Plan to Expel Thousands of Asylum Children;*
**J3** nine samples as follows (dates/titles of articles): 5/1/02 *Theatre of Dreams; Britain's New Role as Force for Good;* 15/10/03 *How On-Demand Can Meet Demands;* 22/10/03 *Flexible Response to Business Challenge;* 12/11/03 *Beware the Weird World of Wireless;* 27/11/03 *Getting the Balance Right;* 28/11/03 *Millions could Die through Flu Vaccine Shortage; Blair Promises Student Fees Rethink;* 29/11/03 *Why Life's Path is Fixed by Three;*

## Magazine subcorpus

**J4** ten samples as follows (dates/titles of articles): 3/1/02 *Astronomers Plan Telescope on Moon;* 4/1/02 *Dolly the Sheep has Arthritis; Euro Coin Accused of Unfair Flipping*; 24/11/03 *Cinnamon Spice Produces Healthier Blood;* 26/11/03 *Squeezed Plastic is Recyclable and Cheap; Predator Theory for Whale Mass Stranding; Blood could Generate Body Repair Kit;* 27/11/03 *Fish Hitch a Ride Upstream on Eddies;* 28/11/03 *Cancer Gene Therapy is First to be Approved*; 30/11/03 *Bear Bones Hint at Osteoporosis Treatment;*
**J5** twelve samples as follows (dates/titles of articles): 5/1/02 *The Semi-Illiterates*; *Give us a Vote Now;* 9/2/02 *Victor's Justice; Life without Death; Misogynist in the Woodpile;* 23/2/02 *Atomic Britain; The Rape of Europe*; 29/11/03 *Trouble in Manhattan; Ancient and Modern*; *Who Hates the Jews Now?; Women who Won't; The Times has gone Tabloid: Where will the Broadsheet Revolution End?;*
**J6** ten samples as follows (dates/titles of articles): 2/1/02 *Fighting the Wrong Battle;* 3/1/02 *Ring in the Euro;* 4/1/02 *A New Way to Fly;* 6/11/03 *Making an End of it;* 13/11/03 *Thanks for No Memory;* 20/11/03 *Thinking about it; A Poxy Story*; 27/11/03 *Unpicking the Fiscal Straitjacket*; 28/11/03 *Trouble on the Fringes; I'd Buy that for a Dollar*;

### *Fiction corpus*

## Children's fiction subcorpus

**F1** two continuous samples: *Cinderella*; *The Sleeping Beauty* (book of fairy-tales), 125-184
**F2** a single continuous sample (novel), 7-25
**F3** seven continuous samples selected at random: *The White Cap; St Dunstan and the Devil; The Ghost of Lady Hobby; St Eustace's Well; Possessed by the Devil; A Beautiful Lady whose Name it was Ruth; Jack o'Legs;* (book of legends), 48-54; 55-58; 84-87; 157-159; 180-183; 196-198; 246-250;

## Adult fiction subcorpus

**F4** a single continuous sample (novel), 3-33
**F5** a single continuous sample (novel), 1-24
**F6** a single continuous sample (novel), 3-18

## 2 External parameters of paragraphs in the corpus

Since in this study the major unit under scrutiny is the graphic paragraph (though in the majority of cases, it coincides with the paragraph as a content unit), it appears worthwhile to examine some of the most conspicuous outer parameters of paragraphing at some length.

### 2.1 Setting off paragraphs

All the Internet sources were marked by the so called American paragraphing (skipping a line to set off a paragraph). These involve the journalistic texts: J1, J2, J3, J4, J5, and J6.

Indentation was characteristic of the rest. No source, however, indented their paragraphs immediately following a headline or a subheadline. The number of the dropped spaces slightly varied. All the sources apart from one indented two spaces, whereas A4 did three.

In most sources, the very initial item was graphically plain. Two sources, however, marked it off by an enlarged initial letter (A3; F1). In a single source, however, we detected capitalization of the first five graphic words (A2).

### 2.2 Quantitative aspects of paragraphing: Two counts

It is perhaps needless to mention that the researcher may theoretically choose to adopt various ways of measuring paragraph length, such as in terms of the number of the words, lines, clauses or utterances enclosed. We selected two complementary approaches, namely measuring paragraph length in terms of utterances and main clauses. The boundaries of the former are signalled very clearly in written language, i.e. by conclusive punctuation marks. The latter approach is obviously much less mechanical. Nevertheless, it also seems to be justified since a main clause is in many ways central to linguistic research. For instance, in FSP it corresponds to the basic distributional field. Furthermore, it is well known that a subordinate clause represents in fact a clause constituent of a C. Naturally, we are fully aware that other approaches to measuring paragraph length could have been adopted instead.

#### 2.2.1 Paragraph length in terms of number of utterances

Examining first the average number of U falling to a paragraph, we notice a striking variability in paragraph lengths, ranging from one to as many as 29 U. It should be noted, however, that paragraphs containing one to three U made up 60.34 per cent of the overall corpus. What is more, paragraphs containing one to

six U amounted to 88.31 per cent of the investigated corpus. It follows that the relatively longest paragraphs, i.e. those ranging from seven to 29 U, were clearly marginalized, covering only 11.69 per cent overall. Interestingly, the extreme length of 29 U was detected only in one source, namely in F5, and, moreover, in a single paragraph. Indeed, very long paragraphs are not perceived as "manageable building blocks".

These findings suggest that paragraph length, as other concepts in linguistics, is rather graded and can be assessed in terms of centrality and marginality. The prototypical end of the cline may be suggestive of the author's tendency to establish a particular paragraphing rhythm, or, put differently, it may be indicative of his/her eye-appeal awareness.

Further we can observe that the frequency of paragraphs, at least in the current research corpus, appears to be inversely related to paragraph lengths. In other words, the longer the paragraph, the fewer cases we find. Thus, in the overall research corpus, the greatest representation was detected with paragraphs containing only a single U (21.74%), followed by two-U paragraphs (20.92%) and three-U ones (17.68%). Further, a four-U paragraph group covered 12.46 per cent a five-U one 8.94 per cent a six-U one 6.57 per cent paragraphs in all. Next, paragraphs enclosing seven U corresponded to 3.57 per cent, those comprising eight U to 2.75 per cent, those involving nine U made up 1.69 per cent of all and, lastly, paragraphs exhibiting ten U formed 1.11 per cent of all. In the overall corpus, groups of paragraphs containing more than ten U were the rarest of all (amounting altogether to less than one per cent of the instances).

Examining individual register corpora separately, we revealed some remarkable discrepancies. The most striking diversity in paragraph lengths was detected in the fiction corpus. The length variability was somewhat less noticeable in the academic corpus and relatively restricted in the journalism corpus. More specifically, in the fiction corpus the longest paragraph had 29 U, in the academic corpus 15 U and in the journalistic corpus eleven U.

The extreme paragraph length varied also across the individual subcorpora. The most striking disproportion was discovered between the two subcorpora of fiction, as the longest paragraph in the children's fiction subcorpus involved ten U, while its adult fiction counterpart as many as 29 U. Presumably, shorter paragraphs were believed to be more motivating and easier to process for younger readers.

The diversity in lengths may be perhaps best illustrated by the number of the length-items (length types), where each item corresponds to a particular number of U enclosed in a paragraph. If we accept the premise that the lower the number of distinct length items in a corpus, the more pronounced the paragraphing rhythm,

we may observe that the greatest diversity was detected in the fiction corpus (22 different paragraph lengths), followed by 14 length items in the academic corpus and eleven length types in the journalism corpus. It appears, then, that in the lastly mentioned corpus, the eye-appeal in paragraphing may play a relatively prominent role.

Turning now to the distinctions between the relevant subcorpora, there were negligible differences between the two academic subcorpora (on average 10.67 length-types per a source in the subcorpus of natural sciences, and 11 length items in that of humanities). Close similarity may be also observed between the newspaper and magazine subcorpora (on average 7.33 length items in the former, and 7.67 length-types in the latter). The most striking differences, however, were detected between the two fiction subcorpora, with seven different length types on average in the children's fiction and 16 length items on average in the adult fiction subcorpus.

The length tendencies become even more apparent if we explore the number of tokens in the subcorpora. We decided to focus on the two extremes of a cline. On the one hand, we studied the number of tokens of the longest paragraphs (exceeding 7 U). On the other, we surveyed the number of single-sentence paragraphs. Paragraphs exceeding seven U turned out to be extremely rare in the newspaper subcorpus (5 paragraphs of 345) and in that of the children's fiction (7 paragraphs). They were more frequent in the magazine subcorpus (35 paragraphs), and in that of natural sciences (48 paragraphs in all). However, they were rather common in the subcorpus of adult fiction (62 instances) and that of humanities (83 specimens).

Conversely, single-sentence paragraphs were most frequent in the newspaper subcorpus (144 cases), and in both the fiction subcorpora (125 specimens in children's fiction and 113 instances in adult fiction). Their occurrence in the remaining subcorpora was somewhat lower, though very even (22 instances in the natural sciences, 23 paragraphs in the humanities and 24 specimens in the magazine subcorpora).

These findings suggest that the paragraphing rhythm was most evident in newspaper texts and children's fiction. Interestingly, both the subcorpora were marked by the brevity of their paragraphs. The natural sciences and humanities texts had longer paragraphs. Adult fiction exhibited the most extreme degree of length diversification with highest tokens in both extremities. Presumably, this makes room for Findra's dichotomy between marked and unmarked paragraphing (1973).

Finally, we explore the average paragraph lengths. The mean of the overall research corpus is 3.60 U per paragraph. Examining individual register corpora,

however, we detect some significant peculiarities: The academic corpus typically features very long paragraphs (4.67 U per paragraph on average). Somewhat shorter are those composing the fiction corpus, displaying on average 3.21 U, whereas the journalism paragraphs tend to be the shortest of all (their mean length corresponding to 2.91 U).

Moreover, there are some striking differences between the subcorpora investigated. The mean value in the humanities subcorpus (5.00 U per paragraph) exceeded that of the natural sciences subcorpus (4.34 U per paragraph). Similarly, the magazine subcorpus displayed on average longer paragraphs (3.75 U) than its newspaper counterpart (2.06 U per paragraph). Finally, with its 4.10 U per paragraph, the adult fiction subcorpus prevailed in paragraph length over that of children's fiction (2.32 U).

Thus we can arrange the subcorpora according to the increasing mean lengths of their paragraphs (measured in terms of the involved U) as follows:

The shortest paragraphs were found in the newspaper subcorpus (2.06 U on average), followed by the subcorpora of children's fiction (2.32 U), magazine sources (3.75 U), adult fiction (4.10 U), and natural sciences (4.34 U), with the humanities subcorpus displaying the longest paragraphs of all (5.00 U).

### 2.2.2 Paragraph length in terms of number of main clauses

An alternative measurement of paragraph length is the number of the enclosed main clauses (C).

Generally speaking, the variability of paragraph lengths in this measurement is even more noticeable, ranging from one to as many as 34 C. Nevertheless, paragraphs containing one to four C made up 61.93 per cent of the overall corpus. Paragraphs featuring one to seven C amounted even to 84.73 per cent of the investigated corpus. In other words, paragraphs ranging from eight to 34 C covered only 15.27 per cent overall. Here again, the extreme length of 34 C was detected only once and in a single source, namely F5.

Similar to the findings pertaining to the number of U embraced in paragraphs, the greater their length (measured this time in terms of the C), the lower the frequency of the respective paragraphs. Thus, the greatest representation may be accorded to paragraphs containing only a single C (17.53%), followed by two-C paragraphs (16.96%) and three-C ones (15.12%). As the paragraphs lengthened, their overall representation gradually decreased. Indeed, a four-C paragraph group covered 12.32 per cent a five-C one 9.57 per cent and a six-C one 7.49 per cent of all. Furthermore, seven-C paragraphs corresponded to 5.75 per cent, eight-C ones to 3.91 per cent, nine-C ones to 3.43 per cent and, lastly, ten-C

ones to 1.84 per cent. Moreover, paragraphs containing eleven to thirteen C amounted to more than a single per cent share each (11 C ones 1.64%, 12-C ones 1.01%, and 13 C-ones 1.06%). In the overall corpus, groups of paragraphs containing over thirteen C were rather marginal (composing less than one per cent of all).

Paragraph length in the corpora showed some major tendencies: Variability was most pronounced in the fiction corpus, less in the academic corpus and least in the journalism corpus. More specifically, the longest paragraph in the fiction corpus had 34 C, in the academic corpus 21 C, and in journalism only thirteen C.

Moreover, the extreme paragraph lengths varied even across the individual subcorpora. Interestingly, the greatest discrepancy was discovered between the two subcorpora of fiction. The longest paragraph in the children's fiction included 15 C, while its adult counterpart comprised as many as 34 C.

The diversity in lengths may be examined also by the number of the length-items, where each item corresponds to a distinct number of C involved in a paragraph. The greatest diversity was found in the fiction corpus (23 different paragraph lengths), followed by the academic corpus (18 different length items) and, lastly, by the journalism corpus (12 distinct length types).

As to the individual subcorpora, there was some increase in length variability between the newspaper and magazine subcorpora (on average eight length items per source in the former, and 9.33 length-types in the latter). Somewhat greater disproportions were found between the two academic subcorpora (on average 12 length-types in samples of natural sciences, and 15 length items in sources dealing with humanities). The most striking differences, however, were detected between the two fiction subcorpora, with ten length-types on average in the children's fiction and 20.33 length items on average in the adult fiction subcorpus.

The length tendencies become even more apparent if we trace the number of tokens. Here again, we decided to scrutinize the two extremes of a cline. First, we focussed on the number of tokens of the longest paragraphs (exceeding 10 C). Second, we surveyed the number of single-C paragraphs. Paragraphs exceeding ten C turned out to be extremely rare in both the journalism subcorpora (one paragraph in the newspaper subcorpus out of 345 and six paragraphs in the magazine subcorpus). They grew in significance in children's fiction and natural sciences (11 instances each). However, the longest paragraphs turned out to be rather common in humanities (48 tokens) and in adult fiction (60 instances).

Conversely, single-C paragraphs were the most common in the newspaper subcorpus explored (115 specimens), and in both the fiction subcorpora (100 examples in adult fiction and 92 in children's fiction). Their incidence in the

remaining subcorpora was rather low (23 specimens in the magazine subcorpus, 17 cases in natural sciences, and 16 instances in humanities).

From these findings it can be inferred that the perceptible paragraphing rhythm measured in terms of the number of the C enclosed in paragraphs is most evident in newspaper sources, followed by the magazine and children's fiction subcorpora. We can observe a striking lengthening between the two academic and fiction subcorpora. Indeed, adult fiction exhibits the top degree of length diversification with the highest figures in the extreme categories.

Finally, we investigate the average length values. The mean paragraph length in the overall research corpus corresponds to 4.49 C. Examining individual style corpora shows some significant tendencies. The academic corpus typically featured rather long paragraphs (5.70 C per paragraph on average). It was followed by the fiction corpus with an average 4.39 C. Finally, journalism paragraphs tended to be the shortest of all (3.37 C).

Moreover, we have noticed some striking differences even among the subcorpora examined. The mean length value of the humanities subcorpus (6.36 C per paragraph) exceeded its natural sciences counterpart (5.04 C). Similarly, the magazine subcorpus displayed on average longer paragraphs (4.31 C) than its newspaper counterpart (2.42 C). Finally, with its 5.29 C to a paragraph, the adult fiction subcorpus prevailed in length over the children's fiction counterpart (3.49 C).

Thus we can arrange the subcorpora according to their increasing mean paragraph length (measured by the number of the involved C) as follows: The shortest paragraphs were found in the journalistic subcorpus (2.42 C on average), followed by the children's fiction (3.49 C), and the magazine (4.31 C) subcorpora. Much longer proved to be the paragraphs in the academic sciences (5.04 C) and the adult fiction (5.29 C) subcorpora, with the academic humanities subcorpus again displaying the top length values (6.36 C).

### 2.2.3 Comparison of the two paragraph length measurements

The tendencies expounded above for the paragraph lengths measured by the number of U and by the number of the enclosed C were found to be strikingly similar. The relative order of subcorpora according to the average paragraph length was shown to be parallel. In fact, there was only a single exception in the two counts, namely, the positions of the adult fiction and natural sciences subcorpora were swapped. The difference suggests that, on average, the U in the subcorpus of natural sciences tend to exhibit a lesser number of C than their adult fiction counterparts.

## 3 Conclusion

Generally, it appears that paragraph length results from an interplay of a whole range of diverse factors. In our analysis we have pursued several of them more systematically. In particular, we have scrutinized paragraphing in various register corpora. We have shown that in terms of utterances as well as in terms of main clauses, the journalism paragraphs were extremely short, academic ones long, the fiction paragraphs in between.

In the register subcorpora the following tendencies were noted: Children's fiction favoured much shorter paragraphs than adult fiction. Newspaper samples were marked by paragraph brevity compared to magazine ones. Natural science texts tended to display shorter paragraphs than their humanities counterparts. It appears, then, that graphic paragraphing is text type/register-specific and that for each of these categories, its norms are set rather differently.

There also seems to be at least some affinity between the length of the paragraphs and the overall length (size) of the sample texts. Indeed, the longest paragraphs were detected in the three subcorpora composed of single-sample sources exclusively, namely in adult fiction (4.10 U/ 5.29 C), and in both the academic subcorpora, i.e. in natural sciences (4.34 U/ 5.04 U) and humanities (5.00 U/ 6.36 C).

Given the fixed amount of paragraphs per a source subcorpus, we could assume that the number of samples composing the source subcorpus is inversely related to the sample's lengths. Hence one could easily hypothesize the shortest paragraphs in sources exhibiting the highest numbers of the enclosed samples. However, as the mere comparison of the values in newspaper and magazine subcorpora convincingly shows, the highest number of samples per a source subcorpus does not automatically produce the shortest paragraphs. It should be recalled that the subcorpus of magazine texts features altogether 30 articles, whereas the newspaper subcorpus only 25 samples. The mean paragraph length, however, is much longer in the former (3.75 U/ 4.31 C) than in the latter (2.06 U/ 2.42 C). Hence it appears that the size of the sample text represents only one of the factors entering the complex interplay.

In what follows, we will attempt a more comprehensive view of the other relevant factors, including those which the above analysis has not yet foregrounded: The author's idiolect may also play a role, as demonstrated here e.g. by Murdoch's longest paragraph of all (i.e. F5). This might occasionally explain even the incidence of notional (i.e. marked) paragraphs (see e.g. Daneš 1994/1995, Dubois 1997).

Another factor, by no means negligible, may be the intended recipient, e.g. the age of the child or adolescent reader; the lay reader or novice to a discipline

(textbook type of academic texts) vs. a peer erudite recipient, etc. This factor affected the mean length differences between paragraphs in source A5 and those in sources A4 and A6.

Moreover, paragraphing is presumably influenced by the subject matter dealt with, especially demonstrable using the two academic subcorpora. It appears that compared to texts dealing with natural sciences, those dealing with humanities call for a remarkable lengthening of paragraphs, owing to their comparably greater emphasis on interpretation, and evaluation, together with their tendency to incorporate various asides, comments, contextual descriptions, and the like.

In addition, regard should be paid to the relevant graphic conventions (e.g. those adopted by certain publishing houses or else, notably those governing paragraphing of direct speech portions). Presumably, general layout factors are also at play. For instance, the narrow newspaper columns turn easily even single utterances into graphic paragraphs. This is presumably why journalists are habitually said to be "one-sentence paragraphers" (see e.g. Baker 1966: 16).

The research into the paragraph-length variation within the register corpora and subcorpora also suggests that a rather prominent role may be accorded to the eye-appeal or paragraphing rhythm, although their impact appears to be most perceptible in relatively closed registers. In the researched corpus, the comparably most pronounced rhythm in paragraphing, presumably governed by the eye-appeal awareness, seems to be characteristic of journalistic writing. Conversely, the greatest variation in paragraph length was characteristic of fiction.

Furthermore, a significant role may be attributed to the type of the selected paragraph build-up, together with the presence (or absence) of a subtle paragraph-internal hierarchy. As we have shown elsewhere (e.g. Pípalová 2005, 2006), higher consistency and stability in paragraph build-up, as a rule, reduce paragraph length. Conversely, build-up instability, inconsistency, or else elaborated internal hierarchy tend to connote lengthier paragraphs. A more thorough scrutiny of this aspect, however, exceeds the focus of the present study.

Last but not least, our analysis corroborates the findings observed for Slovak fiction by Findra (1973). Indeed, with the most striking length variability detected in fiction samples, it seems reasonable to conclude that creative writers occasionally employ marked (notional) paragraphs. Their incidence, which may be perceived only against their more prototypical unmarked counterparts, is linked to the desired stylistic effect, i.e. the intended speeding up or slowing down of the reading pace. Indeed, paragraphing may both, accelerate and retard the motion, and, in this way, underscore the intended dramatic effect.

Thus can be seen as part of the complex writer – reader interaction that deserves much more attention by linguistic researchers. Paragraphing may appear

to be only a formal, "outside" phenomenon, but this analysis has shown that a wide range of factors may influence the conscious or unconscious conventions in the different text-types which this study has analysed empirically.

## 4 Appendix

### MEAN PARAGRAPH LENGTH PER CORPORA AND SUBCORPORA

| CORPORA | Ø Utterances | Ø Main Clauses |
|---|---|---|
| OVERALL CORPUS | 3.60 | 4.49 |
| ACADEMIC CORPUS | 4.67 | 5.70 |
| JOURNALISTIC CORPUS | 2.91 | 3.37 |
| FICTION CORPUS | 3.21 | 4.39 |

| SUBCORPORA | Ø Utterances | Ø Main Clauses |
|---|---|---|
| NATURAL SCIENCES | 4.34 | 5.04 |
| HUMANITIES | 5.00 | 6.36 |
| NEWSPAPER | 2.06 | 2.42 |
| MAGAZINE | 3.75 | 4.31 |
| CHILDREN'S FICTION | 2.32 | 3.49 |
| ADULT FICTION | 4.10 | 5.29 |

**Abbreviations:**

C main clause
U Utterance

**Note**

[1] For the purposes of this study, no further distinction will be made between these terms.

**References:**

Baker, S. (1966) *The Practical Stylist.* 9th ed. New York: Thomas Y. Crowell Company.
Bečka, J.V. (1992) *Česká stylistika.* Praha: Academia.
Brown, G., Yule, G. (1983) *Discourse Analysis.* Cambridge: Cambridge University Press.
Brown, K. (ed.) (2006) *Encyclopedia of Language & Linguistics.* 2nd ed. Oxford: Elsevier Ltd. (12).
Carey, G. V. (1976) *Mind the Stop. A Brief Guide to Punctuation.* Penguin Reference. London: Penguin Books.

Daneš, F., Grepl, M., Hlavsa, Z. (eds) (1987) *Mluvnice češtiny* III. Praha: Academia.
Daneš, F. (1994) 'Odstavec jako centrální jednotka tematicko-kompoziční výstavby textu (na materiále textů výkladových)' (Paragraph as a central unit of the thematic and compositional build-up of the text (On the material of expository texts)). *Slovo a slovesnost*, 55, 1-17.
Daneš, F. (1995) 'Paragraph - a central unit of the thematic and compositional build-up of texts.' In: Wärwik, B., Tanskannen, S. K., Hiltunen, R. (eds.) *Organization in Discourse. Proceedings from the Turku Conference. Anglicana Turkuensia* (14) Turku: University of Turku. 29-40.
Dubois, B. L. (1997) *The Biomedical Discussion Section in Context. Advances in Discourse Processes* XLVI. Greenwich, London: Ablex Publishing Corporation.
Findra, J. (1973) 'Odsek jako prvok epickej štruktúry.' In: *Segmenty a kontext. Litteraria XV.* Bratislava: SAV. 66-115.
Heffernan, J. A. W., Lincoln, J. E. (1982) *Writing. A College Handbook.* New York, London: W. W. Norton & Company, Inc.
Hodges, J. C., Whitten, M. E., Webb, S. S. (1986) *Harbrace College Handbooks.* 10th ed. Harcourt Brace, Jovanovich Publishers.
King, G. (2000) *Punctuation*. Collins Wordpower, Glasgow: Harper Collins Publishers.
Kolin, P. C. (1986) *Successful Writing at Work.* 2nd ed. Lexington, Massachusetts, Toronto: D.C. Heath and Company.
Mathesius, V. (1942/1982) 'Umění psát odstavce. Řeč a sloh.' In: *Jazyk, kultura a slovesnost.* Praha: Odeon. 92-146.
Macpherson, R. (1997) *Základy anglické stylistiky.* Praha: Academia.
Miko, F. (1973) 'Odsek, téma, štýl.' In: *Segmenty a kontext. Litteraria* XV. Bratislava: SAV. 5-29.
Mistrík, J. (1985) *Štylistika.* Bratislava: Slovenské pedagogické nakladateĺstvo.
Miššíková, G. (2003) *Linguistic Stylistics.* Nitra: UKF.
McArthur, T. (ed.) (1992) *The Oxford Companion to the English Language.* Oxford/New York: Oxford University Press.
Pípalová, R. (2005) 'On central paragraphs with a stable P-Theme.' In: Čermák, J., Klégr, A., Malá, M., Šaldová, P. (eds) *Patterns. A Festschrift for Libuše Dušková.* Praha: FFUK, KMF. 157-167.
Pípalová, R. (2006) 'Unfolding a paragraph Theme.' In: Povolná, R., Dontcheva-Navratilova, O. (eds): *Discourse and Interaction 2. Sborník prací Pedagogické fakulty Masarykovy univerzity. Vol. 198*. Brno: Masaryk University. 119-130.
Quirk, R. et al. (1985) *A Comprehensive Grammar of the English Language.* Harlow, London & New York: Longman Group Limited.
Stanford, J. (1993) *Guidelines for Writers. Rhetoric, Reader, Handbook.* New York/St.Louis: McGraw-Hill, Inc.
Troyka, L. Q. (1987) *Simon and Schuster Handbook for Writers.* New Jersey, Englewood Cliffs: Prentice-Hall, Inc.
Warriner, J. E., Whitten, M. E., Griffith, F. (1977) *Warriner's English Grammar and Composition. Third Course.* New York, Chicago, San Francisco, Atlanta, Dallas & London: Harcourt Brace Jovanovich.
Wichmann, A. (2000) *Intonation in Text and Discourse. Beginnings, Middles and Ends.* Harlow: Longman, Pearson Education Ltd.

**Primary Sources:**

**ACADEMIC CORPUS:**

**NATURAL SCIENCES SUBCORPUS**

**A1** Norwich, J. J. (ed.) (1991) *Britain's Heritage.* London: Kingsfisher Books.
**A2** Watkin, D. (1990) *English Architecture. A Concise History.* London: Thames and Hudson, reprinted version.
**A3** Chown, M. (1999) *The Magic Furnace. The Search for the Origin of Atoms.* London: Vintage Random House.

**HUMANITIES SUBCORPUS**

**A4** Clark, J. J. (1994) *Jung and Eastern Thought. A Dialogue with the Orient.* London and New York: Routledge.
**A5** Marsh, S. B., Soulsby, J. (1987) *Outlines of English Law*. 4th ed. London: McGraw-Hill Book Company Ltd.
**A6** Yearley, S. (1991) *The Green Case. A Sociology of Environmental Issues, Arguments and Politics.* London and New York: Routledge.

**JOURNALISTIC CORPUS:**

**NEWSPAPER SUBCORPUS**

**J1** www.guardianco.uk accessed on dates: 4/1/02; 5/1/02; 25/11/03; 28/11/03; 29/11/03;
**J2** www.independent.co.uk accessed on dates: 4/1/02; 5/1/02; 28/11/03; 29/11/03; 30/11/03;
**J3** www.thetimes.co.uk accessed on dates: 5/1/02; 15/10/03; 22/10/03; 12/11/03; 27/11/03; 28/11/03; 29/11/03;

**MAGAZINE SUBCORPUS**

**J4** www.newscientist.com accessed on dates: 3/1/02; 4/1/02; 6/1/02; 24/11/03; 26/11/03; 27/11/03; 28/11/03; 30/11/03;
**J5** www.spectator.co.uk accessed on dates: 5/1/02; 9/2/02; 23/2/02; 29/11/03;
**J6** www.economist.com accessed on dates: 2/1/02; 3/1/02; 4/1/02; 6/11/03; 13/11/03; 20/11/03; 27/11/03; 28/11/03;

**FICTION CORPUS:**

**CHILDREN' S FICTION SUBCORPUS**

**F1** Carruth, J. (1991) *Treasury of Fairy Tales. Goldilocks and the Three Bears.* London: Award Publications.
**F2** Rowling, J.K. (1999) *Harry Potter and the Prisoner of Azkaban.* London: Bloomsbury.
**F3** Marshall, S. (1996) *English Folk Tales*. London: Phoenix Giant.

**ADULT FICTION SUBCORPUS**

**F4** McEwan, I. (1999) *Amsterdam.* London: Vintage, Random House.
**F5** Murdoch, I. (1996) *Jackson's Dilemma.* London: Penguin Books Ltd.
**F6** Tremain, R. (1992) *Sacred Country*. A Sceptre Paperback. London: Hodder & Stoughton.