

Marilena Maniaci

# At the End of the Line: Text Continuity and the Division of Words in Byzantine Manuscripts

From the earliest stages of their development, all horizontally arranged alphabetical writing systems have consistently divided texts into lines of equal length (as far as possible) and set them down in a regular way from left to right or from right to left, with one line placed on top of another until a pre-determined space has been partially or fully occupied.<sup>1</sup> Arrangement of the text in this way constrained

---

Translated from the Italian into English by Mark Livesey. Original published as Maniaci, Marilena (1997), 'Alla fine della riga. Divisione delle parole e continuità del testo nel manoscritto bizantino', in *Scriptorium*, 51: 189–233.

This contribution makes use, in part, of the data, ideas and results presented in a doctoral thesis in Greek and Latin palaeography produced between 1995 and 1998 (*Il libro manoscritto bizantino (secoli XI e XII): supporto, impaginazione, gestione del testo*), which is currently undergoing an in-depth review prior to its publication [note of the editor: published as Maniaci 2002]. I would like to thank Guglielmo Cavallo for his unwavering willingness to follow and to encourage the clear definition of my research interests, and for consistently inspiring me to publish my findings. I would also like to thank Marco Palma, who read a first draft of the text and improved it with his exacting and pertinent observations. My gratitude is also due to Ezio Ornato, who took a lively interest in following the progress of my work and participated in a generous fashion, thereby contributing in a decisive way to the simplification of various problematic issues, but at the same time adding new layers of complexity to others which had previously seemed rather more straightforward to me.

---

<sup>1</sup> It is well known that in order to economise on time and energy, many ancient Greek inscriptions (above all those dating prior to the 5<sup>th</sup> century) were written in a boustrophedonic manner (i.e. 'in the direction of a ploughing ox'), which entails changing the direction of the script at the beginning of each new line, and sometimes also the direction of individual characters, or even—albeit more rarely, and in order to achieve particular aesthetic effects—in a stoichedic style, that is to say one which aligns the letters both horizontally and vertically, as in the boxes of a crossword (see Guarducci 1987, 28–30). It is not unlikely that the abandonment of this peculiar typology of text arrangement influenced the progressive refinement of layout requirements for inscriptions: in fact, boustrophedonic writing necessitated an alternate checking of each of the vertical extremities of the epigraphical rectangle, which entailed a doubling of the effort required on the part of the lapicide or *ordinator*. Conversely, stoichedic writing offered the aesthetic advantage of producing a perfect vertical alignment of the sequence of characters

the medieval scribe to exert a continuous effort in order to conform to a more or less rigid standard of uniformity that respects the grid laid out by himself or others when the page was initially ‘constructed’.<sup>2</sup> On the other hand, the adoption of a unidirectional writing strategy—which requires a not negligible periodic shifting of the eye—is countered by the uninterrupted flow of the text, and consequently hampers the linear nature of the reading process.

The conflict that arises between the continuity of text flow and the discontinuity of the graphic ‘chain’—and, as a result, between segmented writing and uninterrupted reading—manifests itself in a number of ways. The subdivision of the text into unidirectional lines impedes, *ipso facto*, an uninterrupted uptake of the transmitted message. Indeed, every change of line forces the reader’s brain to make an effort to temporarily memorise the spatial coordinates and content of the previous line, at the same time as pinpointing the coordinates of the successive one and verifying the essential coherence of the written message. In the present contribution, this combinational procedure is referred to by the conventional term ‘re-assemblage’.

The success of the ‘re-assemblage’ process—or at least of its comfortable application—is inevitably impeded by a number of factors. Some of these factors are directly linked to the visual perception of the written page: for example, the geometrically uniform succession of complete lines that are seemingly devoid of any individual character, but whose monotony is sometimes mitigated by distinct aesthetic and stylistic choices (typical of certain graphical systems and medieval ornamental devices). Another example is the succession, either intentional or by chance, of alliterations or anaphoras which tend to result in the overlapping of homeoarctons and homeoteleutons, a phenomenon which increases the possibility of faulty ‘re-assemblage’ occurring.<sup>3</sup>

---

at the end of each line, albeit at the cost of a random subdivision of words (i.e. without respecting syllabic division) which happened to fall in this position.

**2** This expression hints at a procedure that consisted in establishing, in advance, the spaces on the surface of the page destined to be occupied by the main text and those to be left blank. This was achieved through the application of criteria and procedures whose after-the-fact reconstruction has proved both difficult and controversial, given the almost total absence of explicit testimonial evidence. See Maniaci 1995; Maniaci / Ornato 1995.

**3** Highly complex—and beyond the scope of my knowledge—is the array of research initiatives aimed at developing interpretative models of the interrelationships existing between the visual structures of writing and the ways in which they are perceived and interpreted through reading. Concerning these matters, which can be regarded as almost entirely modern disciplines, whose methods and approaches, despite being occasionally raised in palaeographical studies, remain largely unknown to manuscript historians as methods and data acquisition approach-

Other factors are dependent on the segmentation of the text into lines, a procedure that interacts in a complex way with the grammatical and syntactical division of the manuscript's contents. In whatever tangible way this 'interference' manifests itself, it exerts an influence—almost always a negative one—on the rapid and accurate reading of a text: if the distribution of the written lines coincides with the natural pauses in the discourse, memorising the content of a line will be made easier, but 're-assemblage' with the successive line cannot be entrusted—in the absence of other indications—to an intuitive reconstruction of the overall logic of the contents. In the opposite scenario, 're-assemblage' will be facilitated by the structure of the phrase, but memorising a truncated message will inevitably create difficulties.<sup>4</sup>

In addition to artificially superimposing itself on the syntactical structure of the discourse, the segmentation of a written text into a series of lines placed one on top of another brings with it the need to repeatedly interrupt the semantic continuity of the message, inasmuch as, assuming one keeps the length of each segment the same or roughly the same (a situation regularly made evident, in manuscripts, by the existence of a 'border line' set down in hardpoint or in colour), the interruption at the end of a line will not always coincide with the end of a

---

es, I will confine myself to directing the interested reader to the thought-provoking review by Mastruzzo 1995, 413–424.

4 It should be noted, in this regard, how common sense, following the logic of verbal language, dictates the rules of typographical aesthetics without too much concern for the functionality of the reading process. Indeed, it is the case that the pauses in verbal discourse have to coincide with syntactical structure: needless to say, the said rules also impose themselves on the reader's 'interior discourse', and punctuation techniques succeed in respecting them perfectly well. However, the segmentation of the text laid out on the page exerts an effect on the reading process at another level, which conforms to a different logic: in this context, the most effective line shift is the one whereby the final word in the previous line most influences the reader's pursuit of the next line. Thus, in the sentence 'Solo l'universo medesimo apparisce immune dallo scadere e languire: perocché se nell'autunno e nel verno si dimostra quasi infermo e vecchio, nondimeno sempre alla stagione nuova ringiovanisce' [G. Leopardi, *Operette morali, Cantico del gallo silvestre*], the seemingly most logical line change occurs after the infinitive 'languire', but instead the most effective one actually falls after the preposition 'dal-lo', inasmuch as a faulty 're-assemblage' with any lexeme other than a substantive (or substantive infinitive)—and only a substantive beginning with an impure 's' or with a 'z'—would raise an 'alarm'. It should be noted that the change after "'l'" or "nell'"—which is also quite effective (since it necessarily connects with a substantive or infinitive beginning with a vowel), inasmuch as it is more open than the previous one—is forbidden by traditional orthographic rules (even if it is in fact increasingly tolerated). Also greatly frowned upon in the world of typography is a change of line between a reference sign and the reference itself: 'see p./348' upsets the most sensitive reader, even if it provides an effective reconnection with the successive line.

word.<sup>5</sup> The division of words at the end of a line (from hereon in referred to by the term ‘splitting’<sup>6</sup> for the sake of simplicity) therefore constitutes a practically inevitable phenomenon which is intrinsic to the process of transcribing texts—and not only in the Greek context, which constitutes, as we shall soon see, the chief subject of the present research.<sup>7</sup>

Precisely on account of its widespread diffusion, in addition to the repetitiveness with which it occurs—and consequently the ease with which it can be observed—the phenomenon of word splitting offers a highly propitious area of study when conducting research into the legibility of the medieval book.<sup>8</sup> On the other

---

5 Needless to say, it is always possible to conclude all the lines of a text with a whole word, on condition that one is prepared to forgo the regularity of the two side margins, which appears to have represented, for texts in prose, an ideal that was already very widespread in the practice of manual text transcription. Curiously, in modern times this ideal had to be temporarily renounced when, with the advent of the typewriter, the inflexibility of the machine’s mechanism made it too difficult to calculate the spaces that had to be left between one word and another in order to achieve perfect justification of a line. Not by chance, the automatic justification of lines was one of the greatest advantages vaunted by word processing systems when they were first launched on to the market.

6 The not altogether felicitous expression is used here in the absence of a better one (and also to avoid resorting to even more bold neologisms, albeit possibly more accurate, such as ‘lexitomy’ (Italian = ‘lessitomia’), which is equivalent to the French ‘coupure des mots’ and the German ‘Worttrennung’). The lack of an unambiguous, specific and established term to define, in its various aspects, the phenomenon under discussion provides an indication of the scant attention which has been paid to it up until now (not only in the field of palaeography, but also in the fields of orthography and the psychology of reading), a neglect which justifies the introduction of neologisms, with the aim of avoiding the need to continually resort to cumbersome circumlocutions.

7 The same can confidently be affirmed with respect to Latin and the principal modern languages of Indo-European derivation. It can also safely be presumed that other linguistic systems had to address similar concerns.

8 The specific problem of determining, after-the-fact, the legibility of medieval manuscripts has seldom been addressed—and then only at a theoretical level—in the palaeographical bibliography. In order to obtain a general impression of the set of problems concerned, one should consult, in particular, Bergeron / Ornato 1990, as well as some of the results obtained from an investigation of a sample of Latin and French codices dating from the 15<sup>th</sup> century presented in Bozzolo et al. 1987, 130–133. As regards the Latin context, the relationship between the changes seen in textual presentation—i.e. the separation of words, punctuation, and other aids to the reader—and the evolution of reading methods is elucidated above all in works by Paul Saenger, Malcolm Parkes and Richard and Mary Rouse (to which essential references can be found in the bibliographical review cited at the end of this footnote; additional titles are listed in successive footnotes). The dialectical debate on the copying process and the reading process during different eras of the Latin medieval period, along with reflections on the presentation of texts are addressed in Petrucci 1984. The interested reader can also consult the anthology edited by

hand, the analysis of strategies employed by copyists to optimise the segmentation of texts represents just one aspect of a much larger set of problems—a set of problems that we can identify with the blanket expression ‘line management’, namely the combination of principles that guided the copyist and the stratagems that he put to practical use so as to create pages that were at one and the same time both legible and aesthetically harmonious.<sup>9</sup>

In order to justify the choice of a phenomenon which, on the face of it, appears to be somewhat marginal as a topic for research, as a first step it is necessary to frame it within the wider context of a ‘history of legibility’ (yet to be written in its entirety) and a ‘history of copying by hand’, whose mechanisms have only been partially understood and outlined.<sup>10</sup> With this objective in mind, one can draw a parallel between the written page and a ‘mine field’ that conceals potential ‘reading slip-ups’ which occur quite frequently, and are fairly serious in nature—slip-ups whose frequency and gravity prove, in all cases, to be considerably increased when the reader is also the copyist, and therefore reads with the immediate aim of transcribing. This is because (a) any errors made in the copyist’s work might be perpetuated in successive copies, and (b) because the ‘re-assemblage’ process is far more tiring—and therefore riskier—for an amanuensis who is obliged to perform visual ‘gymnastics’ by constantly shifting his gaze between his copy and the model he is working from, than it is for an ordinary reader. In other words, the process of transcription involves something above and beyond simply reproducing a text in a new volume without losing any information, whilst at the same time respecting a set of fundamental aesthetic rules that more or less resemble those seen in the model. Indeed, the work of transcription entails tackling a series of tangible problems whose resolution depends not only on the objective difficulty of applying specific rules,<sup>11</sup> but also—to a large extent—on the ability of

---

Cavallo / Chartier (eds) 1995, with a detailed and well-thought-out bibliography on pp. 441-461. Unfortunately, only when the writing of this article had already been completed did I become aware of the contribution by Supino Martini 1996. On p. 40, with respect to the Latin world, the essay also touches on the set of problems that concerns us here, although somewhat oddly no mention is made of the previous methodological reflections of Bergeron / Ornato 1990, which pay particular attention to this issue.

**9** Needless to say, both of these criteria are subjective, and therefore susceptible to variations in time and space: indeed, a text can be adjudged to be more or less legible, based on the competences of the person reading it—or on those of the audience/person destined to read it—or more or less aesthetically pleasing in relation to codified customs and individual tastes.

**10** A chapter on the ‘copyist at work’—providing essentially the same information—can be found in all the principal textbooks on palaeography, codicology and textual criticism.

**11** Needless to say, the problem increases in relation to the number and inflexibility of the constrictions that the copyist—by his own volition or that of others—has to tackle. It is obvious,

the individual scribe. Therefore, if the scribe performs his task in an intelligent way, transcription also means assisting the reader (with stratagems that will vary in relation to his familiarity with the written material) by foreseeing and seeking to resolve in advance any ambiguous situations. However, the act of foreseeing necessitates work which, depending on the circumstances—i.e. the ways and extent to which stylistic traditions, functional needs and technological rigidity exert their influence—can prove to be quite demanding. For this reason, it is likely that the solutions of the said technical problems will take multifarious forms and will vary according to the different geographical, cultural and historical contexts concerned.

But in what way, in more concrete terms, can the division of words at the end of lines compromise the smooth progress of the reading process?<sup>12</sup> It is obvious that the presence of split words will never totally prejudice the correct interpretation of a text: indeed, the phenomenon can be compared to the formation of small whirlpools in a homogeneous body of water which impede its flow to a certain extent, but at any rate only temporarily. If we consider the simplest case—one which, moreover, is well-represented in the transcription of medieval texts—where a part of a word is ‘forwarded’ to the successive line without the use of a hyphen, it is already easy to see how merely the theoretical possibilities alone that this event raises might represent a source of continual ambiguity for the reader who, once he has reached the end of a line, will always find himself in one of the following two situations, briefly summarised as follows:

- (a) The final lexeme of the line (Z) cannot be considered semantically independent (in other words, it does not make sense in the language of the text).
- (b) The final lexeme appears to be semantically independent.

In case ‘a’ it is probable that the Z is the first part of a split word, but on the other hand it is not entirely out of the question that one is dealing with an error

---

on the other hand, that the material ‘translation’ of a text into a new form will prove to be far more demanding as the difference in presentation (i.e. size, text arrangement, script, decorative elements, abbreviations, etc.) between the model and the copy to be made increases.

<sup>12</sup> Here, the expression ‘reading process’ is intended to mean the personal way in which a text is absorbed. Abstracting from this case—in many ways a rather particular one—the reading process contributes to the copying process. However, it should be noted in this connection (where the possibility of comparing an antigraph with one or more of its direct apographs exists) how textual criticism can serve as a very useful tool for the in-depth study of the ‘history of legibility’.

made by the copyist (or a typographical error), or alternatively an uncommon word that the reader is not familiar with.

On the other hand, in case 'b' it is probable that the Z is in fact a whole word, but one might also be dealing with the first half of a split word which, owing to the freakishness of the case, has been afforded semantic independence.

In both cases, an examination of the first lexeme of the successive line (A) has to be carried out and its semantic value duly assessed. This necessitates determining not only the semantic value of Z and A considered separately (Z+A), but also that of Z and A conjoined (ZA). However, the recomposition of a semantically valid sequence is not sufficient: it is also necessary to confirm whether or not it proves to be compatible with the context of the discourse. With this objective in mind, the penultimate word of the line (Y) is examined, together with the second word of the successive line (B), in order to verify the respective pertinence of the groups Y+Z+A+B and Y+ZA+B. Even in cases where the reader is well versed in this type of exercise, the necessity to carry out all the verifications can only slow down his reading speed and hamper the 're-assemblage' process. In addition (obviously) to the personal attitude of the reader, the relative ease and speed of mental verification depend on the context, which is to say the degree of likelihood that the splitting of a word at the end of a line might create a misunderstanding, and the degree of plausibility of the misunderstanding generated.<sup>13</sup>

Regarding these general observations on the division of words at the end of lines, one could raise the objection that such examples of potential ambiguity are seldom encountered in the real world. In fact, the reality is precisely the opposite, even if, naturally, extreme<sup>14</sup> cases (H) are quite rare. For example, one has only to

---

<sup>13</sup> Thus, for example, *non potest parvo res magna constare* would not create any difficulty; *in ea coniu ratione apprehendistis...* all things considered, would be less confusing than *illi | furi | bundo...*—where the first part of the split word, which has a full meaning, agrees with *illi*—in that the semantic value of *coniu* attenuates, in advance, the effects of the misunderstanding raised by the *ratione apprehendistis* in the following line; *mea est par | vitas illis...*—in which the three initial lexemes follow a logical course and do not immediately appear to be incompatible with the grammatical flow—would probably create greater confusion, but would seem less ambiguous than *mea est par | vis pauperibus...*, in which only the irrelevance of the context would represent a source of concern. Fortunately, phrases such as *sine virtute vir tute non vivit...* do not crop up very often in Latin.

<sup>14</sup> In actual fact, a statistical analysis carried out on an incunabulum of good quality in the Italian language (Sallustio, *Opera*, Venezia, 1470, Windelino da Spira) demonstrated that more than 40% of cases in which the first or second fragment of the split word at the end of a line are endowed with semantic independence, and that 7% of both cases, considered separately, have a meaning in the language that the text is written in (which does not necessarily mean, obviously, that their succession is grammatically and syntactically coherent). (The relevant data

remind oneself of the fact that in inflected languages the division of words all too easily results in grammatical ambiguities (e.g. *matri-bus*; *amavi-mus*; λύω-μεν; whilst the abundance of prefixes frequently represents the source of modifications, or even semantic inversions such as *au-fero* | *af-fero*; ἀνα-βαίνω | ἐκ-βαίνω.<sup>15</sup>

A second objection, on the other hand, would not cast doubt on the ways in which the phenomenon appears, but instead would question the extent of their objective effects and their subjective impact. In other words, the reading difficulties caused by the division of words could be judged to be broadly speaking insignificant—or at least, not sufficiently serious so as to raise much concern on the part of medieval copyists and readers—above all in contexts where the division of words is not yet fully apparent and established in the graphic chain.<sup>16</sup>

This kind of objection can be countered by two striking facts: first of all, there is the use in some manuscripts, be they in verse or in prose—and of various content—of so-called subscripted letters (Italian = ‘codini’, that is ‘little tails’). This involves the final letters of the final word on a page being transcribed into the

---

were kindly supplied by Ezio Ornato, who I duly thank.) Even in the absence of analogous statistics, one can assume that the Greek language is less susceptible to ambiguities of this kind, if one bears in mind its systematic accenting notation and the phonetic phenomena arising from the meeting of suffixes and desinences, and from apophony, all of which can modify the presentation of the theme and the position of the accent during the inflection.

**15** It should be noted that in such cases the division of words respects the lemmatical and etymological structure of the word involved. Such a circumstance—which on first glance would seem to assist the reader in comprehending the text—turns out in reality to be detrimental to a smooth and correct reading.

**16** Here, it should be recalled that in the Byzantine minuscule the coincidence between graphic unity and semantic unity took place gradually over a number of centuries and was never fully accomplished. Up until a time which has not yet been clearly identified through specific studies (but at any rate, after the 12<sup>th</sup> century), the distribution of blank spaces responded—broadly speaking—to a logic imposed by the capacity of individual letters to ‘hook up’ (towards the right) with the succeeding letter. The rules governing such a logic should be studied from diachronic and synchronic standpoints (some general observations—and corresponding research perspectives—on the tracing of the Byzantine minuscule can be read in Canart 1990. Additional bibliographical references are provided *ivi*, footnote 2). As regards the Latin world, Paul Saenger placed the widespread adoption of the ‘canonically separate’ script at the peak of a process that took place between the 11<sup>th</sup> and 12<sup>th</sup> centuries; this was the product of a demand for expository clarity by the medieval schoolmen and was closely related to the shift from reading aloud to silent reading (see, among the numerous contributions dedicated by the author to this subject, Saenger 1982; Saenger 1990; Saenger 1995). Starting at the beginning of the 12<sup>th</sup> century, the first to abandon *scriptio continua* in the transcription of Latin, thereby forcing themselves to isolate one word from another with the help of new punctuation marks, were the Irish scribes (see Parker 1987; Parker 1992, 20-29; Parker 1995).



lower margin located below the final line, so as to avoid having to transfer the entire word to the following page.<sup>17</sup> From the stylistic point of view, ‘codini’ are always segments of split words, inasmuch as the word concerned is divided into two fragments. However, the second fragment, instead of being inserted into the textual flow of the successive line, is transcribed under the preceding one, and therefore retains a specific and unambiguous connection with it. A solution of this kind has the advantage of eliminating any ambiguity in relation to the value of the final lexeme of the line, but at the same time it has the disadvantage of spoiling the uniform appearance of the page’s presentation. It is therefore understandable that this solution is only occasionally encountered—and is limited to particular situations—in medieval manuscripts.<sup>18</sup>

The case with respect to hyphens<sup>19</sup> is different, since these were destined to be widely used both in handwriting and in modern typography. If one considers the simplicity and immediacy of the use of this sign, it is quite surprising to note how it only succeeded in establishing itself rather gradually, and even then, not fully, during the course of the Middle Ages.<sup>20</sup> This is true not only in the sense that hy-

---

**17** The term ‘codino’ is derived, by analogy, from the typographical lexicon, in which it is used to refer to one or more lines that, during the composition of a text, exceed the length of the writing area, but which cannot be transferred to the following page. Such fragments are therefore placed in the lower margin, a small distance away from the text, in order to be ‘re-absorbed’ (see Fioravanti 1993, 132, and Maniaci 1996 [1998<sup>2</sup>], 167). Less frequently, in Byzantine manuscripts one also encounters ‘codini’ formed not from a few letters, but instead from the final words of a sentence (see, for example, the codex Vat. gr. 268, 11<sup>th</sup> century, in which about half a line often appears in the lower margin). Scant precise knowledge is available on the spread of the phenomenon and the circumstances that govern the presence of such fragments of text; however, one notes, for example, that in Greek two-column manuscripts ‘codini’ generally appear only under the second column, and very seldom—understandably—under the first column, although the text can easily extend from under one column to under the next, remaining on the same page. On the other hand, ‘codini’ that coincide with lines other than the last, similar to those encountered in Hebrew manuscripts (although these serve a different purpose), are entirely unknown (see below, footnote 40).

**18** As will become clearer in due course, the use of ‘codini’ does not follow any entirely regular and coherent pattern. See below, footnote 119.

**19** Maniaci 1996 (1998<sup>2</sup>), 205.

**20** The situation appears to be less clear in the case of the printed book, inasmuch as in manual typesetting, thanks to the total inflexibility of the writing area, the systematic division of words, marked by the presence of the hyphen, necessarily requires additional work. There is a lack of precise data on the precise moment in time when the hyphen appeared and spread in Byzantine manuscripts, but in any event its use remains altogether exceptional throughout the entire 12<sup>th</sup> century. Even if the phenomenon has not yet been the subject of any systematic survey, I can at least state that in our corpus one meets with very few instances: see, for exam-

phens do not appear in all codices, but also in that within one and the same codex—and even, quite often, on the same page—they can either be present or absent, proving at worst to be more troublesome than helpful. It is not, in fact, difficult to understand that if the presence of a sign is positively associated with a phenomenon, its absence will be regarded negatively, for which reason the sporadic application of a rule inevitably results in a false sense of certainty, and is therefore ultimately more detrimental than if there were no rule at all. Here, one is dealing, as in many other instances, with an example of the remarkable degree of sloppiness that characterises medieval manuscript production—a sloppiness which, to our eyes, appears to be present even when one cannot cite, as a partial justification, the rudimentary nature of preindustrial tools. It is a behaviour that is perhaps more ostensible than real, but in any event the ‘key’ to understanding it continues to evade us.

Whatever the case may be, the existence of subscripted letters and hyphens demonstrate that the problems raised by the division of words were very real, and as such were certainly experienced by medieval readers. It is also clear that whatever his technical ability may have been, the scribe was never entirely unconcerned about line change; indeed, the rule that insists on respecting the syllabic structure of words when they are divided established itself very early on, in the Greek and Latin worlds, at least. The existence of such a rule—which even in the absence of precise statistics can be considered to have been widely applied<sup>21</sup>—obliged the scribe to be capable of foreseeing the need to divide a word, as well as being able to decide on the precise point at which to make the division.

---

ple, the manuscripts Vat. Barb. gr. 525 [10<sup>th</sup>–11<sup>th</sup> centuries]; Vat. Barb. gr. 521 [11<sup>th</sup>–12<sup>th</sup> centuries]; and Vat. Pal. gr. 203 [11<sup>th</sup> century]). Only at a very advanced stage of the preparation of this essay did I become aware of Ruffini 1996, which I was not able to consult directly. Today, the development of ever more powerful and sophisticated word-processing programs that are able to minutely adjust spacing between letters and words (i.e. microjustification) so as to achieve the highest possible degree of line filling, has once again led to the redundancy of word splitting at the end of lines, and also, as a result, of the use of hyphens. In Latin manuscripts in Caroline minuscule of Italian origin the hyphen seems to disappear starting from the third/forth decade of the 11<sup>th</sup> century, according to indications in Petrucci 1968. According to Saenger 1990a, 453, the ‘invention’ of the hyphen is owed to scribes working in Benedictine monasteries in southern England in the first half of the 8<sup>th</sup> century, who later on introduced its use to scriptoria of continental Europe.

<sup>21</sup> The rule had already been codified by grammarians in the Early Middle Ages: Petitmengin 1985, 107–108 and footnote 113, cites the case of Bishop Victor of Capua, who between 546 and 547 revised the orthography of the ‘Codex Fuldensis’ (Fulda, Landesbibliothek, Bonif. 1, CLA 1196) and corrected the division of words falling at the end of lines, based on rules inspired by Greek grammarians.

## 1 The necessity to split words: a simulation experiment

Subscripted letters and hyphens overcome the more serious problems arising from the division of words, but not the actual phenomenon itself: whichever remedy is applied, it is obvious that an intact word will always be more legible than one which is divided in two. Remedies demonstrate a clear desire to reduce the awkwardness resulting from the division of words, but not a concern to eliminate them, as far as possible, at their root. Theoretically, the two attitudes are not incompatible, but neither are they equivalent. One cannot, in fact, exclude the possibility that the total absence of split words was seen as the ideal goal in some instances, and that the remedies were used in situations where the ideal was simply not achievable.

How, then, can one ‘measure’ the existence of a working method aimed at preventing the appearance of the phenomenon one wishes to study—in other words, the absence of a phenomenon? Merely counting the number of lines in which the integrity of the final word is preserved is not sufficient, since whilst the presence of a split word always has a negative connotation—i.e. chance dictated that the final word on a line was too long to be included in its entirety—the significance of its absence is somewhat ambiguous: either fate had it that the length of the final word was not excessive, or the writer<sup>22</sup> ensured that the length of the final word coincided with that of the final segment in the available line.

A better and more scrupulous investigative approach—and one which can be accomplished in a reasonable space of time—consists in counting the number of lines that terminate with a divided word and then comparing the result of the count (carried out on a sufficiently large number of lines) with that obtained from a transcription selected entirely at random. In other words, a random probability model is employed which serves as a basis for drawing comparisons with observed reality.

Reducing the problem to its bare essentials, let us picture an alphabet whose letters are of invariable width, on a page whose lines, which likewise are of invariable length, are circumscribed by a strictly delineated border and contain  $N$  characters. Such a scenario, despite being simplified to the maximum, can be observed in a real situation (even if this is one which is now

---

<sup>22</sup> Here, the term refers as much to the copyist (scribe) as to the typographical compositor.

rapidly approaching extinction), namely that of the typewriter.<sup>23</sup> In the case of a typewriter, a line is totally without flexibility; indeed, the transcription process moves inexorably from the beginning to the end of each line, and the writer cannot resort to the use of any expedient that will afford him the least amount of leeway to operate.<sup>24</sup>

Now, let us imagine an ideal ‘raffle box’ containing all the words making up the language being used, *weighted according to the frequency with which they occur*<sup>25</sup>, together with a mechanism that successively extracts a certain number of them and compares, upon each extraction, the number of characters  $C$  already positioned on the line with the amount of space  $S$  remaining, which is to say with the number of characters that the line can still accommodate. The process is repeated until the value of  $C$  is lower than the value of  $S$ . If, after the final extraction, it is found that  $C = S$ , the mechanism will register a ‘success’ and go on to fill the next line; if, on the other hand, the result is  $C > S$ , a ‘failure’ is registered, meaning that the final word has to be split.<sup>26</sup>

What then, in theory, is the probability, respectively, of each outcome occurring? In order to be able to calculate this in a precise way, one would have to know the distribution of the word lengths—expressed through the number of characters they contain—in the language of the text undergoing analysis. The probability of word splitting will in fact be greater as the average word length increases, and in particular as the frequency of long words increases.<sup>27</sup>

---

**23** The adoption of a proportional alphabet, which constituted the norm in manual and mechanical working practice in medieval times, complicates matters, as we shall see later on. This does not, however, greatly change the overall nature of the problem.

**24** Such flexibility corresponds, in essence, to the degree of tolerance that one can afford oneself with respect to the pre-established number of characters per line in order to avoid splitting the final word.

**25** This means that the imaginary ‘raffle box’ would include a higher number of the most common words, and a proportionately lower number of the rarest.

**26** If we ignore the rather cumbersome presence of an opponent who conceals his cards, our imaginary mechanism can be likened to a baccarat player, or a player of the more popular card game ‘Seven and a Half’, who has to get as close as possible to attaining the maximum score, and ideally to achieve it. At the same time, though, he must not exceed the maximum, otherwise he will be disqualified. In any event, the player, even if he completely discounts the value of the successive card, enjoys a certain advantage, since he knows in advance the composition of the pack, and is therefore able to calculate, if he takes into account the sum total of the points already scored, the chances of his ‘going bust’.

**27** In the case of a proportional alphabet—where the words *fili* and *mamma*, which are both composed of five letters, but which are not remotely of the same length in terms of the amount of space they occupy on a line—it would be necessary to know the length of individual words in millimetres (or typographical points). In other words, the length expressed as the number of

However, even if the distribution of long words were consistent, for practical reasons such a calculation would still be impossible to carry out.<sup>28</sup>

In such circumstances one inevitably has to resort to a solution which, despite being far more simplified than the real-life situation, permits the analyst to create a good approximation of the theoretical model. This consists in simulating, with the aid of a computer, the conditions encountered during a real-life text transcription, by using a typewriter whose ‘terminal margin’ (i.e. tab stop)<sup>29</sup> has been blocked. For this purpose, two passages are inserted into a spreadsheet<sup>30</sup> (at the rate of one word per rectangular field, without considering punctuation),<sup>31</sup> one in Latin, one in Greek, with each text composed of 550 words.<sup>32</sup> Next, for each pas-

---

characters has to ‘weighted’ with respect to the width of each of the characters. In this way, the simulation would inevitably become more complicated, but the fundamental problem would remain the same. On the length of words in Latin and Greek, see footnote 35 below.

**28** If one considers, for example, a line whose length is fixed at 50 characters (like the one that will be adopted later on in our simulation), the difficulty consists not so much in determining the splitting probability for each  $n < 50$  value hypothesised from the length of the text (this will be provided, obviously, by the frequency of all the words whose length is greater than  $50 - n$ ), but rather in establishing the probability of each  $n < 50$  value, in relation to which the need to carry out a word split division is assessed. In fact, as soon as the value of  $n$  becomes quite large, the number of possibilities to achieve it through a given sequence of words becomes astronomical ( $n = 6$  can already theoretically equate to: 6 words of 1 letter in length; 1 word of 1 and 1 of 5 letters; 1 of 2 and 1 of 4 letters; 2 of 3 letters; 3 of 2 letters; and 1 of 3, one of 2 and 1 of 1 letter, and so on, in 17 possible permutations, each with different probabilities. Combinatorial analysis (which forms part of the general theory of probabilities) does not provide a mathematical formula to solve such a problem.

**29** This neologism is preferable to the term ‘right margin’, which, needless to say, is valid for writing systems that progress from the right to the left.

**30** A spreadsheet is a form of software that permits the management of data inserted into lines and columns (which intersect each other to form a network of rectangular fields) that makes it possible—through the application of appropriate formulae—to perform calculations (including complex ones), apply statistical tests, create tables and generate two- and three-dimensional graphs of various kinds, all a lot more easily and rapidly than by using manual procedures.

**31** Naturally, given that one is dealing with transcriptions of modern printed editions, the samples do not envisage the presence of abbreviations and acronyms. Clearly, this does not mean (as we shall see) that an assessment of the latter has no relevance within the framework of the set of problems being addressed. See below, 635 ff.

**32** The pre-selected Latin text is the opening of the Life of Romulus (*de Romulo primo Romanorum rege*) from *De viris illustribus di Francesco Petrarca* (Martellotti [ed.] 1964, 6-8); the Greek text is drawn from Plutarch, *Caesar*, 63-64 (Ziegler [ed.] 1968, 327-330). It is only natural that both the textual typologies and the datings of the texts are not without consequence with respect to the distribution and the length of the words they contain. To reduce potential distortions as much as possible, two texts of similar content were selected.

sage a page composed of fifty ‘pseudo-lines’ is created, with each line containing eleven words, and all of them (obviously) containing a variable number of characters. Within each ‘pseudo-line’ formed in this way, a ‘real line’, limited to fifty characters, is created—ideally by advancing the carriage of the typewriter—which, depending on the length of the words it accommodates, either entails or does not entail a final word division.<sup>33</sup>

The test was performed under conditions of increasing ‘tolerance’. To begin with, the number of 0 tolerance splits was calculated: this situation implies that the scribe was neither able to stop writing before he reached the margin nor to exceed it. In order to avoid starting a fresh line by splitting the final word, it was therefore necessary for the fiftieth letter to coincide *exactly* with the end of the word. Successive tests were carried out with tolerances of  $\pm 1$ ,  $\pm 2$ ,  $\pm 3$  and  $\pm 4$  characters.

Needless to say, the results obtained by applying such a test are inevitably subject to random variations, the extent of which it is worthwhile to verify by subjecting a sufficiently large number of text samples to the same procedure. Given that the inclusion on the calculation sheet of several dozen passages would have significantly hampered the practical side of the operation, the variability of the samples was also simulated by randomly ‘shuffling’ the words and lines.<sup>34</sup>

---

**33** In theory, for the Latin sample, to the fifty characters envisaged per line, ten characters equating to the spaces between words should have been added. It is easy to see how the introduction of such blank spaces, being constant in number, would have pointlessly complicated the experiment, and furthermore would not have exerted an effect on the outcome. For the Greek writing, in which the distribution of the blank spaces never coincides exactly (in the centuries concerned) with the distinction between words (see above, footnote 16), the problem is only *seemingly* more complex: in the splitting of words, the scribe in any event referred to semantic unity, and did not allow himself to be swayed by their physical segmentation on the page. However, this does not mean to say that this particular distribution of blank spaces will not have repercussions for the problem we are presently addressing, inasmuch as even the blank spaces within individual words could be used by scribes to adapt the text to the length of the line.

**34** Concerning the simulation carried out on the different passages, the approach settled on raises two problems, namely the population of words is always the same, but on the other hand, the sequence obtained is no longer a ‘text’, if by this term we mean a succession of words that form a logically coherent whole, rather than merely a series of words existing separately from each other in the language concerned. However, provided that the distribution of the length of words in the pre-selected sample reflects in a sufficiently faithful way that of the population of origin, the shuffling will result in a massive number of possible combinations, which guarantees sufficient variability. As regards the coherence of the text, this aspect does not appear to exert a negative effect on variability: indeed, it is easy to grasp how the probability of a word division occurring depends in the first place on the length of the final word that can be accommodated by the line, and in the second by the length of the sum of the preceding words. The latter parameter is not dependent on the order in which the words appear. On the contrary, one might say that the variability of the samples could be

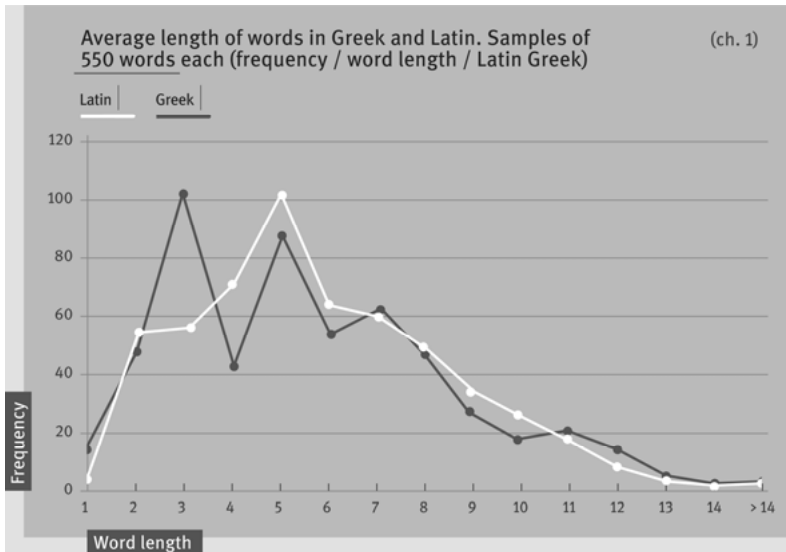
The samples composed in this way more than anything make it possible to identify the parallels between Greek and Latin with respect to the length of words. Such parallels are not immediately obvious in advance, since the two languages—both of which are inflected—differ, owing to the presence in Greek of an article, which in theory should reduce the average length of words. In reality, the overall average word length is practically the same in both languages (5.49 for Greek, and 5.80 for Latin),<sup>35</sup> which reveals, when viewed from this perspective, a large degree of structural similarity, even if distribution trends in each language do not coincide perfectly (Chart 1).<sup>36</sup> The similarity observed with respect to the length of words is reflected in the theoretical distribution of word splitting, which remains almost the same in both languages, regardless of variations in the tolerance level (Tab. 1).

---

even greater in an incoherent discourse, inasmuch as a succession of very long or very short words is perfectly compatible with the rules of the particular case, but this is more unlikely to happen in reality.

**35** The Greek sample effectively contains a manifestly higher number of words composed of 1, 2 and 3 letters (the bulk of them being prepositions and articles—the latter not existing in Latin), but Latin is richer in words in the next band up, composed of 4, 5 and 6 letters, which to a certain extent compensates for the variation between the two language systems. The greater presence of short words in the Greek language is also reflected in the wider scattering of the distribution (the coefficient of variation—which is to say the relationship between the quadratic deviation and the mean—equates to 53.3% for Greek and 46% for Latin. One can offer as an example the count carried out on two samples composed of 550 words, with respect to the frequency of words in relation to the number of syllables. If the words are arranged in order based on an increasing number of component syllables (from 1 to 7), the words included in each group turn out to be, respectively, 137, 200, 128, 96, 1, 1 for Latin, and 191, 144, 130, 57, 22, 5, 1 for Greek. One notes that, as expected, Greek contains more words composed of one syllable (mostly articles), whereas Latin is richer in disyllabic words. This difference notwithstanding, the sum of words which contain from one to three syllables is identical in both languages. It should also be noted that the two means with respect to the length of words used in the text represent only an approximation vis-à-vis the medieval transcription process employed for Greek and Latin texts, owing to the effect exerted by the different rates of abbreviation employed in the two separate linguistic contexts (which are difficult to compare in the absence of systematic studies for the various periods concerned), and by the width of the omitted letters (for example, it is well known that the letter most frequently eliminated in Latin is M, though it is also the widest letter, whilst Greek minuscule script does not include abbreviations for π, which together with ω, is usually the widest letter in that language).

**36** It is interesting to observe that a sample composed of just 550 words is sufficient to produce curves that are well defined and clearly convergent, in contrast to what happens with other phenomena related to the grammatical and syntactical characteristics of texts (i.e. the frequency of determinant words or constructions), which result in much wider variations (e.g. discrimination in the case of lexical analysis and textual criticism, for attribution and stylistic characterisation purposes).



**Chart 1:** Average length of words in Greek and Latin. Samples of 550 words each (frequency / word length / Latin Greek)

Theoretical distribution of word splits in a Greek and a Latin sample of 550 words (tab. 1)

'Elasticity'	Greek text	Latin text
0	73.2	69.0
±1	41.4	43.2
±2	22.2	22.8
±3	9.6	10.4
±4	2.6	3.3

**Tab. 1:** Theoretical distribution of word splits in a Greek and a Latin sample of 550 words

As should be plain, the frequency with which word division is resorted to rapidly diminishes with the increase in tolerance, until it touches on zero. The explanation for this phenomenon is clear enough: the computer works in a rather 'dim-witted' way, albeit with utter efficiency. This means that, if the margin of toler-



ance is  $\pm 4$ , all the words of fewer than nine<sup>37</sup> letters in length escape being divided. It does not mean, however, that all the words of more than nine letters in length will necessarily be split, inasmuch as this depends on the sum of the lengths of the preceding words; indeed, the probability of word splitting is positively correlated to this parameter.<sup>38</sup>

In the case of a line entirely lacking any degree of tolerance, the average frequency of word splitting is about 70%. Can we permit ourselves to use this result as an aleatory ‘yardstick’ with which to conduct an investigation of the working practices employed by medieval copyists, then? The answer is in the negative: the copyist, even if he wanted stubbornly to forego the possibility of exercising any control over the issue which concerns us here, could not have worked within a margin of error of zero, because the rule that demands respect for the syllabic structure of a word cannot be applied if one does not allow oneself a certain degree of approximation in either one direction or the other. Unfortunately, we do not have at our disposal any statistics on the distribution of the lengths of syllables in various languages, but we can quite reasonably assume that long syllables are in a clear minority both in Greek and in Latin.<sup>39</sup> Working within a margin of error of  $\pm 1$ —in other words, being prepared to accept that the last word of a line can terminate with a character lying either before or after the page justification—therefore makes it possible to respect the rule in the great majority of cases. Under such conditions, the average frequency of word division descends to under 42%; this is the value that will be adopted for the present investigation as a basis for comparison.

To minimise the need to split words at the end of a line therefore necessitates endowing it with a certain degree of ‘elasticity’. Here, the term ‘elasticity’ is intended to mean the capacity to ensure that the end of a line coincides with the end of a whole word. The easiest way to achieve this aim—which is the same, ideally, as that used in our experiment—consists in simply ignoring, up to a point, the

---

**37** Here and elsewhere, the spaces between words have not been taken into account. The presence or absence of the same, as has already been stated (see above, footnote 33), does not have any effect on the theoretical modelling of the process.

**38** Taking up again the previously expressed formula, and continuing on the basis of the  $\pm 4$  hypothesis, if  $C \cdot n - 1 = 45$  (where  $C$  stands for the number of characters per page and  $n - 1$  the penultimate ‘draw’), only words composed of more than 9 letters will be split. However, if  $C \cdot n - 1 = 40$ , only words composed of more than 14 letters will be divided.

**39** In the two samples consisting of 550 words, the number of syllables ranging from 1 to 5 letters is, respectively, 104 (1), 761 (2), 445 (3), 52 (4), 1 (5) for Latin, and 119 (1), 546 (2), 530 (3), 75 (4), 3 (5) for Greek. Syllables composed of 1–3 letters prevail in both languages, even if Latin has more syllables composed of two letters, and Greek more syllables composed of three.

‘terminal margin’, and instead concentrating on either a lower tolerance (i.e. ‘containment’) or a higher one (i.e. ‘overrun’). In such circumstances the margin will act as a simple ‘signal’, whose function can be compared to that of the margin bell on an old mechanical typewriter. This solution can be applied immediately and does not require any special attention to be paid on the part of the writer. On the other hand, it also breaches to a greater extent the regularity of the writing area’s boundary.<sup>40</sup> For this reason—apart from in cases of low-grade transcriptions, and especially those intended for personal use—the copyist never has enough freedom to drastically reduce, and without too much trouble, his need to resort to word division. Nevertheless, in most cases it will be possible for him to eschew, up to a point, respect for the margin line, and to resort—albeit without going too far—to the two antipodal possibilities, namely overrunning the margin or starting a fresh line before reaching it, with the dual aim of respecting syllabic rules at the same time as avoiding breaks in the shortest words.<sup>41</sup>

In all instances where the copyist either cannot or does not want to ‘abuse’ the elasticity of the line—by adapting its length to a given sequence of whole words—he will, on the contrary, be forced to adapt the word sequence to the length of the line. In theory, this goal can be accomplished by making one or more of the following elements more ‘elastic’:

1. The tracing of letters. The writer can either compress or expand the script in a horizontal direction, based on the available space—albeit at the cost of sacrificing the regularity of the graphic chain, and therefore the appearance of the page.<sup>42</sup>

---

**40** The arrangement of the text ‘flush right’ or ‘flush left’ conflicts with the existence of a visibly traced delineated writing area on the page, and therefore violates the ‘regularity principle’ which represents the guiding principle behind many aspects (as far as possible) in the making of the medieval book. See Ornato 1994, 9. In Eastern cursive scripts in the Arabic and Hebrew world, a concern for respecting the outer margin is made evident in the use of slanting script—positioned slightly below the line—in the final word, in order to avoid an overrun. See Beit-Arié 1992, 38 and Fig. 9.

**41** For prototypographical editions, on the other hand, the border formed by the terminal margin is no longer merely a psychological limit, but instead forms a real barrier that necessarily has to be reached but not exceeded. Viewed from one standpoint, the copyist’s task was more demanding, but viewed from another it should be noted that he enjoyed the ‘luxury’ of working on a line of ‘virtual’ text which could be ‘dissolved’ or incorporated—in theory, at least—at will, up until the moment the work was printed.

**42** In reality, the aesthetic disadvantages resulting from excessive spacing out or compression of script can be reduced by using briefer allographs at the end of lines (although this is only possible with a few letters), or alternatively by superimposing letters. Conversely, a typographer is totally unable to intervene on the width of individual characters but, in addition to

2. The text. The writer can vary the length of the ‘actual’<sup>43</sup> text according to the space available by introducing any necessary abbreviations to adapt the length of a word to the space in which it has to be accommodated, or, vice versa, by foregoing (with the same goal in mind) the most common abbreviations.
3. The ‘non-text’, meaning the spaces between words, the size of which can vary according to specific needs. This expedient, which in theory can also be applied in handwriting, is the one best suited to typographical applications, for which it is today universally employed.<sup>44</sup>

To verify whether or not, and to what degree (and with what objectives in mind) one or the other expedient was employed in the production of a medieval book would call for a minute examination, line-by-line, of a vast number of pages in a sufficiently large sample of volumes. Whatever his interest may be at the outset, it would be unwise for a researcher to undertake such a laborious investigation without first confirming the validity of his initial postulate, namely whether or not the copyists of a given period and work environment were to some extent aware of the problems that we earlier on attributed to word divisions, and whether or not they actively sought to apply remedies whenever possible. This is precisely the

---

employing allographs, he can, if necessary, expand the space between one character and another (which is to say the unprinted space, known as ‘spacing’ in typographical jargon). The optimal distance between one letter and another is known as ‘kerning’. In manual typesetting this space can be expanded as required, whilst its reduction is limited by the inflexible nature of lead type. Conversely, in photocomposition, kerning—also known as ‘compression’—can be reduced even up to the point of superimposing letters. See Fioravanti 1993, 436–437 and 5.

**43** Here, the word ‘actual’ is intended to mean the number of characters present in the text as it was transcribed, in contrast to the ‘virtual’ text, which contains all the characters that are theoretically necessary for the transcription of the same text in a specific alphabet, without resorting to abbreviations. Needless to say, the same ‘virtual’ text can correspond to a very high number of different ‘actual’ texts.

**44** Due to shifting attitudes and continual technological development, the application of the regularity principle to the graphic presentation of the book gradually became more exacting; thus, in order to adapt the text to the length of a line, today’s typographer is no longer able to use the full range of solutions that his medieval predecessor had at his disposal. Nevertheless, the regularity of the page layout—thanks to a carefully calculated and uniform distribution of spacing on lines—is today assured by the speed and efficiency of the automation offered by electronic typesetting techniques. However, manual typesetting (linotype) already made it possible to distribute blank spaces equally—without having to resort to complex calculations or great artisanal skill—through the use of a system of metal wedges which, when inserted into the blank spaces, ‘stretch out’ a composed line so that it reaches the two outer limits of the justification (Fioravanti 1993, 283).

hypothesis that, through the research that follows, we shall attempt to verify by focusing our attention, for now, on the Byzantine manuscript.

## 2 Recourse to word division in a corpus of Byzantine manuscripts

The following results were obtained from a sample of 700 Byzantine parchment manuscripts whose dates range from the 9<sup>th</sup> to the 14<sup>th</sup> centuries.<sup>45</sup> In addition to some other characteristics that will be introduced in due course, the word divisions observed on 15 lines of three non-consecutive pages situated at the beginning, middle and end of each volume respectively, for a total of 45 lines, were counted.<sup>46</sup>

The average percentage of word splits observed across the entire corpus is slightly under 40% (i.e. 37.6%), which is a little lower than that observed in a simulation carried out on 550 words of Greek sample text (41.4%).<sup>47</sup> Therefore, the question has been asked whether this apparently slight variation is purely the prod-

---

**45** More precisely, the manuscripts are divided according to each of the centuries considered: the 9<sup>th</sup> century 23; the 10<sup>th</sup> century 192; the 11<sup>th</sup> century 280; the 12<sup>th</sup> century 167; the 13<sup>th</sup> and 14<sup>th</sup> centuries 38. Naturally, the observations put forward in the text are valid for the three most representative centuries. Furthermore, it should be made clear that the fact that the sample does not include paper codices does not appear (at first glance, at least) to be relevant to the issue under examination. The choice simply depends on the fact that the manuscripts were gathered and examined in the context of a wider research initiative which was dedicated more generally to the construction and utilisation of pages in Byzantine manuscripts on parchment. For making it possible for me to carry out, in a relatively short time, the examination of a large number of volumes, I gratefully acknowledge the generosity of the Prefect of the Vatican Apostolic Library, Father Leonard E. Boyle, and of the Vice Prefect Monsignor Paul Canart. I also gratefully acknowledge the help offered by the personnel working in the library's Manuscripts Department. For the assistance provided in assembling a suitable sample, I must thank Monsignor Canart in particular for making available, with characteristic generosity, his personal catalogue.

**46** Obviously, one is dealing with a highly simplified surveying protocol, given the rather large size of the sample involved and the desire to contain within reasonable time limits this preliminary survey. A far more detailed (and consequently far more arduous) survey to gather the necessary data in order to carry out a more in-depth examination of the phenomenon has been suggested by Bergeron / Ornato 1990, 182–188.

**47** See above 611–613 and footnote 32. This is the percentage obtained if one allows for a 'reasonable' degree of elasticity, equal to  $\pm 1$  character.

uct of chance, or instead if it represents a consistent and sufficiently significant difference so as to justify extending the analysis.

A statistical approach makes it possible to resolve the two opposing hypotheses by comparing the data harvested from the corpus with those we can theoretically expect to obtain—according to the calculus of probability—from a ‘binominal distribution’ (Tab. 2).<sup>48</sup>

Real and simulated number of word splits (based on their theoretical frequency) in the Byzantine corpus		
Number of split words per 45 lines	Actual number	Theoretical number
0	1	0
6 (13%)	11	0
12 (27%)	92	17
18 (40%)	296	301
24 (53%)	258	349
30 (67%)	19	32
36 (80%)	18	0
45 (100%)	0	0

**Tab. 2:** Real and simulated number of word splits (based on their theoretical frequency) in the Byzantine corpus

In fact, the examination of 45 lines of writing in a manuscript can be equated to an  $N$  series of ‘repeated trials’<sup>49</sup> in which the number of instances—in our case split words—taken into consideration is counted. Such instances are commonly known as ‘successes’. If  $P$  indicates the probability of a suc-

<sup>48</sup> Concerning the ‘binominal distribution’ or ‘Bernoulli’s distribution’, see Blalock 1984, 191 onwards, and Giusti 1990, 448 onwards. In order for the function to be applicable, it is necessary that the probability of obtaining a success be the same for each test, and that each test be independent from all the others, and also that each experiment be repeated a predetermined number of times. Our sample, which consisted of  $45 \times 700$  identical observations in individual volumes (different codices and, in the vast majority of cases, ones transcribed by different copyists), fulfilled all the necessary prerequisites.

<sup>49</sup> The classic example employed in textbooks is that of coin tossing (i.e. the probability of getting a certain number of ‘heads’ and a certain number of ‘tails’ when a coin is tossed  $n$  times). See Blalock 1984, 152 onwards.

cess, in virtue of the function known a ‘binominal distribution’ one can calculate the probability  $P$  of attaining, in an  $N$  series of trials 0, 1, 2, ...  $n$  successes; in other words, in our case, the probability of attaining 0, 1, 2, ...  $n$  word splits in a sample composed of 45 lines, if we suppose  $P = 0.41$ , which corresponds to the average number of splits met with in a random experiment.<sup>50</sup> Once 700 observations have been made, the product of  $p \times 700$  will provide the theoretical frequency of the 0, 1, 2, ...  $n$  splits within our corpus, a figure which can then be compared to the instances actually observed.<sup>51</sup>

With respect to the theoretical distribution, the manuscripts included in our sample are more heavily concentrated than expected in the band that corresponds to the lowest percentage of split words (under 27%), which reveals, overall, the existence of a desire to avoid word divisions at the end of lines as much as possible. It remains to be shown whether or not such an ideal was uniformly distributed—in other words, whether or not the majority of copyists tended, in general, to avoid splits (albeit without making an inordinate effort to do so), or alternatively whether the problem was noticed by only a certain number of them who subsequently paid more attention to it.

To provide an answer to this question we can once again apply the test carried out above, although its basic parameters have to be changed, because the basis for comparison in this instance takes the form of a random binominal distribution where the frequency of word splits does not coincide with the theoretical one, but instead with that actually observed in the corpus (Tab. 3).

---

**50** In fact, the mathematical law known as the ‘Law of Large Numbers’ (or LLN) shows that the average value observed in a sample, or in a series of samples, constitutes the best undistorted approximation of the probability value  $P$ , and that the more samples considered, the closer to the expected value it becomes. See Giusti 1990, 415.

**51** In cases where the differences between the theoretical frequencies and the empirically observed ones appear to be of little relevance, the  $\chi^2$  test permits one to verify their statistical significance. (The  $\chi^2$  test makes it possible to compare an empirically observed distribution with a theoretical one, or to compare two empirically observed distributions. The higher the value is, the lower the probability will be that the difference is attributable purely to chance. Results can vary according to a distribution calculated by statisticians. Benchmark values are presented in specially produced tables, depending on the margin of error—or probability—that one is prepared to accept, which in current practice is fixed at 5%. It should be noted that the  $\chi^2$  value is very sensitive to the size of the sample involved; for small samples, it is difficult to obtain meaningful results. See Blalock 1984, 349 onwards).

Real and simulated number of word splits (based on the observed frequency) in the Byzantine corpus

tab. 3

Number of split words per 45 lines	Actual number	Theoretical number/ actual average (38.11%)
0	1	0
6 (13%)	11	0
12 (27%)	92	52
18 (40%)	296	413
24 (53%)	258	226
30 (67%)	19	9
36 (80%)	18	0
42 (93%)	0	0

**Tab. 3:** Real and simulated number of word splits (based on the observed frequency) in the Byzantine corpus

Once again, a comparison with the theoretical distribution shows that the volumes containing only a few split words are more numerous than one might expect. However, a similar phenomenon can be observed at the far end of the distribution, namely in codices that exhibit a large number of split words (which are more numerous in the actual corpus than they ought to be in theory), whilst on the other hand, for the manuscripts that fall within the confidence interval (between 27% and 40% of split words in all the lines), the value for the actual corpus is somewhat lower than it is in the theoretical one. If the data is presented on a graph, the actual occurrence of word splits in the volumes examined would appear to be more ‘compressed’ than in the theoretical one, or in other words, it is less concentrated around the mid values. The dispersion observed at the two extremities of the actual distribution range probably reflects the ‘mixed’ composition of the corpus, i.e. the merging of two distinct groups of volumes that correspond to different attitudes on the part of the scribes with respect to the splitting of words at the end of lines. In one group, it would seem that the copyists placed a great emphasis on limiting the number of divided words, maintaining them under an overall average of 37.6%,<sup>52</sup> whilst in the other it would seem that the copyists

<sup>52</sup> Needless to say, this purely indicative value does not imply that scribes specifically counted the maximum number of word divisions they were prepared to accept on each page of text. In fact, it

were entirely unconcerned by this contingency and tolerated a rate of word splitting which is very close to that of the theoretical average revealed in the simulation experiment (41.2%).<sup>53</sup>

### 3 Factors that influence recourse to word division: an initial hypothesis

In order to confirm the validity of the hypothesis formulated above, it is necessary to establish whether or not the cause of word splitting can be attributed to one or more characteristics in the codices being examined. This necessitates carrying out a systematic survey of the entire sample in order to identify some basic variables. However, an investigation aimed at discovering the reason (or reasons) lying behind the phenomenon requires that a few precautions be taken.

First of all, it is not enough to point to a systematic connection between the number of split words and another parameter of any kind, because the latter might automatically be assumed to explain the phenomenon. Indeed, many of the codices' characteristics are interdependent upon each other, and therefore it is essential to ascertain that the effect exerted by one of them, even if it is a very real one, is not merely a concomitant factor, or simply the product of another more fundamental characteristic.<sup>54</sup> For example, possible variations seemingly linked to the number of lines per page will probably reflect the influence exerted by a volume's size (i.e. the sum of its width and height)—upon which to a large extent this

---

should not be forgotten that the average numbers of splits calculated by us in 45 lines of each manuscript conceal fairly significant differences among the three samples composed of 15 lines.

**53** The hypothesis concerning the simultaneous presence of two distinct populations seems to be contradicted by the finding that the distribution of actual volumes is characterised by a 'peak' (mode) between 40% and 53% of word splits (see Tab. 3). This 'peak'—which, furthermore, is more pronounced than that seen in the theoretical distribution—can be explained by the fact that it very likely combines the maximum values of one of the two populations and the minimum values of the other. One runs the risk of this illusory effect arising each time the averages of the two populations are quite similar to each other and the spread attributed to the classes is too wide to permit an accurate analysis of the distribution trend.

**54** In borderline cases, a seemingly significant effect can turn out to be entirely illusory when subjected to a more detailed analysis. As is well known, the results of pre-election polls can be used in a misleading way if one does not take into account the sociological composition of the sample to which they relate, a step which is indispensable in order to obtain an accurate and correct interpretation of the collected data. The eliminable 'structural effects'—although not always easy to recognise and draw attention to—represent the main foe of any statistical investigation.



parameter depends—and will disappear as soon as the corpus is further subdivided according to the dimensions of the volumes it contains.<sup>55</sup>

Second, it is obvious that any phenomenon is easier to study as a general trend if one focuses one's attention on its most conspicuous manifestations—in other words, if the analysis is limited to the two extreme ends of the distribution, leaving out the central portion. This is particularly true for phenomena seen in manuscript volumes, where each and every example represents the fruit of an artisanal activity in which the craftsman's level of freedom—no matter how his task was organised and regulated—is always quite wide. For this reason, based on the rate of word division, the corpus under examination has been divided into three classes, namely volumes which exhibit a rate of less than 30% (under the average); volumes which exhibit a rate of 30%-42%; and finally those which show a rate of above 42% (above the average rate). By temporarily setting aside the central group, which is of more heterogeneous composition, it becomes possible to examine in greater detail two far more clearly differentiated groups of practically the same numerosity. One group is composed of 152 volumes in which the number of split words is too low not to be the result of a deliberate effort to limit its occurrence, whilst the other, contrarily, is composed of 166 volumes in which the word splitting occurrences are sufficiently numerous to enable us to rule out the existence of any serious effort made to limit them.

A relatively simple method to confirm the existence of a link between the phenomenon in question and any other variable consists in the subdivision of the latter into two 'opposite' classes in relation to an average value,<sup>56</sup> and then to

---

<sup>55</sup> Strictly speaking, the effective validity of all bipartitions that appear to produce meaningful results should be verified by crossing the value that seemingly determines it with the highest possible number of other variables. In our case, in order to simplify the procedure, the variables that can be assumed to lack any connection with the phenomenon in question (for example, ruling systems or leaf signatures) can be excluded. Owing to the inevitable subjectivity of such value judgements, this procedure carries the risk of overlooking a certain number of relevant crossings.

<sup>56</sup> It is important to remember that the variables which can be subjected to statistical analysis are of three types, as follows: qualitative variables that can be organised in nominal groups, and which cannot be placed in anything other than an arbitrary order, and therefore can only be judged on the basis of their similarity or dissimilarity (e.g. the position of quire signatures, either on the writing area or in the margin); qualitative variables that can be organised in a graduated order, that is arranged in an ascending or descending series, but which cannot undergo arithmetic processing (i.e. 'bad', 'mediocre', 'good' or 'outstanding' parchment quality); and qualitative variables that can be organised in a graduated order, and which can all be subject to arithmetic calculations (see Giusti 1990, 49–51). Whilst for qualitative variables the most obvious approach (indeed, the one adopted for this research) consists in taking the 'average value' to be the arithme-

cross reference them with the two extreme ends of the split word distribution. In this way, one can build a ‘contingency table’<sup>57</sup> composed of four boxes (as shown in the example of Tab. 4). It can be said that the cross reference is significant or ‘effective’ if it is able to alter the subdivision of the two original groups of word splits (Tab. 4a)—which we have already seen to be almost equal<sup>58</sup>—and ‘ineffective’ if the opposite is true (Tab. 4b). It need hardly be said that in the real world one never encounters a situation that corresponds to one of the two possible scenarios described; rather, one typically encounters an intermediate situation, where the tendency towards one or the other extreme has to be evaluated on a case by case basis by applying the  $\chi^2$  statistical test.

Contingency table: model of distribution of word splits according to another variable				tab. 4
	a		b	
	split words -	split words +	split words -	split words +
Variable -	80	80	Variable -	160   0
Variable +	80	80	Variable +	0   160

**Tab. 4:** Contingency table: model of distribution of word splits according to another variable

In order to identify the factor—or factors—which held the potential to influence the attitude of Byzantine scribes vis-à-vis the division of words at the end of lines, three types of characteristics were examined. In the first place, aspects attributable, broadly speaking, to the general concept of the ‘quality of the codex’ have to be taken into consideration. It is obvious that such a concept, which is difficult to define in objective terms, is not expressible in a satisfactory way by means of an indicator which is at one and the same time concise and devoid of any ambiguity.

tic mean of the distribution (even if in the case of very unbalanced distributions it is a good idea to utilise a different intermediate value so as to obtain two acceptable classes), in the case of ordinal qualitative variables all the gradations are reduced to just two classes that incorporate several sub-groups. In both cases, the classes obtained will be identified in the present text using the name of the variable followed by a + or - sign.

**57** The ‘contingency table’ provides the simplest graphic representation of the interdependence of the two variables, with one inserted in lines and the other in columns. See Giusti 1990, 122.

**58** 47.8% as opposed to 52.2%. In Tab. 4, in order to simplify matters, it was hypothetically supposed that the sample is cleanly divided into two groups, each being composed of 160 manuscripts.

Therefore, one must settle for considering a few concrete characteristics—which are unquestionably linked to the cost of raw materials and the manufacturing and transcription time—that combine to determine, in more ways than one, the quality of a codex. These characteristics include the quality of the parchment,<sup>59</sup> the presence or absence of gilding, the complexity of ruling,<sup>60</sup> the exploitation of available space<sup>61</sup>, and the regularity of the script.<sup>62</sup>

Another entirely plausible *a priori* hypothesis is that the mechanism which determines the splitting of words depends to a certain extent on the graphic typology of a given manuscript and, within each typology, on the greater or lesser degree of speed with which the script was executed<sup>63</sup>. The inherent difficulties associated with the classification of Byzantine scripts—and, more generally speaking, the ‘reduction to pure numbers’ of the characteristics of any handwritten script<sup>64</sup>—prevent us from being able to subject this particular facet of the prob-

---

**59** Subdivided into four categories, and grouped progressively as follows: ‘poor’, ‘mediocre’, ‘good’, and ‘outstanding’.

**60** Expressed by the total number of surplus marginal lines, subtracting all those which are indispensable for the delineation of the writing area (including the lines that define the columns of manuscripts with ‘double justification’, which it was deemed unnecessary to consider separately, on account of their uniform distribution throughout the sample).

**61** This is quantified by two parameters: the so-called ‘black’—i.e. the relationship between the written area and the total area of the page, which serves as a page filling gauge (Bozzolo et al. 1984, 195–221, and 203)—and the average area of parchment occupied by an individual character (expressed in mm<sup>2</sup>), a concise indicator which is dependent on the size of characters, the height of the line and the overall dimensions of the page.

**62** Expressed by the sum of the coefficient of variation of the width and height of the letter ‘o’. Even if this parameter can reflect the influence of systematic variability factors—voluntary and involuntary modifications, either recurrent or non-recurrent, of the script module (Bischoff 1996)—it depends, to a large extent, on the transcription’s level of accuracy.

**63** If the existence of an interaction between writing speed and the frequency of word divisions at the end of lines appears, at first glance, to be very likely, it is rather less easy to establish whether or not this translates in concrete terms into an increase in, or containment of, the number of split words. On the one hand, the greater level of self-assuredness that typifies a swift hand might seem rather incompatible with the need to anticipate the length of final words in order to be able to avoid having to divide them; on the other hand, a fluid and agile ductus might lead to inadvertently exceeding the bounds of the justification, and therefore make it possible to avoid the need to split the final word on a line.

**64** The potential—and limitations—of the use of a statistical approach in palaeographic studies, as well as the problems associated with establishing suitable parameters for the morphological measurement of writing speed and fluidity, certainly merit investigation, not only in theoretical terms but also on the basis of results gained from actual research. An example of the information that can be obtained by carrying out a series of basic measurements can be

lem to a detailed verification process. Nevertheless, it was still considered worthwhile to carry out a rudimentary survey<sup>65</sup>, based on an examination of various purely dimensional aspects of the script (i.e. the relationship between the width and height of the nucleus<sup>66</sup> of a character; the relationship between the height of the nucleus and the total reach of the ascending stem; the difference between the total maximum height of the letters that can be written within a quadrilinear system and the ruled line; and finally the relationship between the height of the ‘nucleus’ of a character and the ruled line).<sup>67</sup>

Third, it was considered necessary to determine whether or not a link exists between the quantity of split words and several variables that define the relationship between the script and a line. On the one hand, this has to do with the average width of a character (*larcar*)—which equates to the product of the length of the line divided by the average number of characters it accommodates—and the relationship between that value and the length of the line, expressed by the average number of characters accommodated by a full line of script (*medcar*)<sup>68</sup>. On the other hand, it also has to do with the attitude adopted by the copyists with respect to the ‘terminal margin’, which appears on the page as the right justification (this

---

found in Bischoff 1996, which utilises the ‘temporal series’ method to furnish an accurate analysis of an 11<sup>th</sup>-century Latin scribe’s ‘writing rhythm’.

**65** For this purpose, data relating to some basic variables were used. The said data were obtained during previous research projects—or ones that are currently underway—that examine the construction and utilisation of the page in Byzantine manuscripts.

**66** This is the modular relationship in the narrowest sense (calculated by using the letter omicron as a basis). For the present study, this was considered preferable to the modular relationship applied in the so-called ‘Lobbes Lectionary’ (Bruxelles, Bibliothèque royale 18018) by Gilissen 1973, 20–32. As is well known, in this parameter—which represents an abstraction obtained by relating the average width of letters to the average height of their nuclei—factors beyond the control and capabilities of the scribe converge, such as the frequency with which individual letters occur or, more indirectly, the number and nature of abbreviations (see Orna to 1975, and Bischoff 1996, already mentioned above).

**67** Despite the lack of clear morphological characteristics, it is reasonable to suppose that by associating these parameters it should become possible to distinguish among different graphic styles, albeit in an approximate way, but in any event well enough to make it possible to carry out a preliminary statistical analysis.

**68** Clearly, for two-column manuscripts, instead of the length of the two adjacent columns without textual continuity, set out using only one guideline, or alternatively two neighbouring guidelines (i.e. ‘physical line’), the length of a single column (i.e. ‘logical line’) along which the textual sequence unfolds, was considered. As regards the counting of the number of characters per page, this was carried out on ten lines of text (i.e. twenty ‘logical lines’, in the case of two-column volumes), without taking into account blank spaces.

is an attitude that can be evaluated, and measured, by means of the previously mentioned 'overrun' phenomenon).

The first two variables are related to the space that the script occupies horizontally. In what way, then, can this exert an impact on the number of split words occurring? It is a widely held belief that the splitting of a word at the end of a line becomes harder to avoid when the space available on a line is restricted (space being expressed by the number of characters that a line can accommodate). In reality, this belief holds true only in a situation where vertical justification at the open end of the line takes place by intervening, either mechanically or manually, on the width of the spaces between words which have already been composed.<sup>69</sup> In all the other cases the necessity for, and probability of, word division depend purely on the length of the final word—or alternatively on the final sequence of consecutive letters, in the case of scripts where words are not strictly separated.

However, both the average number of characters on a line and the average width of the script could have exerted a psychological effect on the copyist in two different ways: if a line accommodates only a few characters, the probability of having to resort to splitting a word is already greatly increased even when the 'terminal margin' is still quite a long way off, and therefore a copyist who wished to avoid splitting a word had to be alert to the problem earlier on and act with greater assiduity. If the script module is large—and therefore harder to 'compress' without creating a displeasing visual effect—the copyist might then fear that the space available to him will turn out not to be large enough to accommodate the final word, and therefore he would opt for word splitting even in cases where he could have avoided doing so.<sup>70</sup>

With respect to 'overrunning'—which is measured by the number of lines where the script exceeds, even if only by a small amount, the right-hand justifica-

---

**69** In such circumstances, the 'elasticity' of the line is proportionate with the number of blank spaces—and therefore with the number of words—that it contains. If the space that has to be added in order to achieve a perfectly aligned justification and the number of words in the line is limited, the words would seem excessively spaced out, and therefore it would be necessary to resort to word division so as to avoid creating a displeasing effect.

**70** It should be noted that the two parameters 'average number of characters per line' (*medcar*) and 'average character width' (*larcar*) are correlated, but not equivalent. The attitude adopted by a copyist might have been different if he had had to arrange characters of equal breadth on a short line, which would contain fewer characters, or on a longer line, which would contain a greater number. Needless to say, in the former case the impact of the two contributing factors will of course be cumulative.

tion<sup>71</sup>—in theory this is a phenomenon which is inversely correlated to the number of split words, and can either represent a convenient way of intentionally limiting divisions, or instead it can betray rather careless behaviour (on the part of the copyist) which was not expressly intended to fulfil this aim.

It will not escape the reader's attention that the set of characteristics which can be associated, *a priori*, with the splitting phenomenon is still missing an important component, namely the frequency of abbreviations. Indeed, it is easy to understand the importance of these in relation to the 'management of lines' for the scribe who sought to avoid resorting to word division at the end of lines, while at the same time trying to limit the number of 'overruns' beyond the right-hand justification. Unfortunately, the abbreviation count—which relates to a sample composed of 60 words—was carried out on only the first part of the corpus and was reduced to its simplest expression, namely the overall number of abbreviations present in a given number of words.<sup>72</sup> It need hardly be said that in such circumstances the variable cannot be exploited in a systematic way, nor with the degree of complexity that the problem demands.<sup>73</sup>

---

71 In the case of manuscripts with 'double justification'—which represent almost half of our corpus (with respect to the frequency of their occurrence in the Byzantine context, see Maniaci / Ornato 1995, 180–182)—'overrun' was measured twice, using as a reference point the innermost vertical or that which is most distant from the narrow column. In fact, it should be obvious that the latter case, despite being rarer than the former, is a lot more significant. In hindsight, the criteria adopted for counting the occurrence of 'overruns' proved to be too rigid; indeed, it would have been better to consider only cases where the justification boundary was exceeded not by merely one stroke—a very frequent occurrence—but instead by at least an entire character.

72 The counting of abbreviations only involved the 386 manuscripts dating from the 11<sup>th</sup> and 12<sup>th</sup> centuries which were the subject of the doctoral thesis cited above, footnote on page 599. For the remainder of the sample, it was necessary to forego carrying out similar counts, owing to the large amount of time that would have been necessary to accomplish the task. The conflict arising from scientific demands and the amount of time necessary to gather sufficient data—which represents an inevitable problem when it comes to research based on very large collections of material—can only be resolved if multiple researchers collaborate on the same project. For obvious reasons, such collaborations are not always easy to organise or to accomplish.

73 For example, it is obvious that the abbreviations observed in the first and last words on lines should be placed in and counted in two separate categories. In this way, any systematic and consistent imbalance in favour of the abbreviations occurring at the end of lines will provide proof that the copyist employed them as devices to limit the need to resort to splitting words.

## 4 Word splitting and text layout

All the spatial partitions in relation to the quality of the codex proved to have no effect on the distribution of the number of split words. This ‘non-result’ does not have the same value as a simple negative finding. Rather, it demonstrates that in the world of the Byzantine book—at least during the centuries covered by the present investigation—limitation of the number of split words was not explicitly included among the criteria which had to be fulfilled in order to manufacture a codex of good quality.

The same lack of a correlation characterises the partitioning linked to the way in which writing was laid down, which would seem to imply—contrary to one’s expectations—that attention to word splitting neither depended on graphic typology nor on the faster or slower speed with which the writing was executed. In other words, limiting split words was not included among the characteristics associated with a given script or writing level.

Equally, the absence of any connection between the frequency with which splits occur and the number of occasions that ‘overrunning’ is resorted to is quite surprising. In fact, one would normally expect word splitting to occur significantly less frequently<sup>74</sup> in manuscripts where ‘overrunning’ is seen very often. However, this is not in fact the case—not even in manuscripts with double justification where the script tends to exceed even the boundary set by the outermost vertical ruling line.<sup>75</sup>

---

<sup>74</sup> It should be noted, however, that there is a negative correlation between the number of split words and the variation in the average number of characters per line (*CVmedcar*) measured using the coefficients of variation (= variation in relation to the mean) of the average number of characters in a line. However, the correlation becomes noticeable only in codices in which the average number of characters per line is for the most part variable (*CVmedcar* higher than 12%), which is only the case in about twenty volumes. In other words, *CVmedcar* is an indirect measurement of ‘overrunning’, since it indirectly quantifies the degree of irregularity (i.e. ‘indentedness’) of the right-hand justification—or, in other words, the amount of ‘elasticity’ permitted by the scribe vis-à-vis the length of a line. Naturally, this is a rather approximate indicator, and one which should be substituted by a direct count of instances of ‘overrun’ and containment observed in a given number of lines, so as to avoid distortions generated by the random distribution of narrow or wide letters of the alphabet on the lines considered. The incidence of this last-mentioned factor obviously increases in relation to the proportionality of the alphabet—or in other words, in relation to the differences in the width of narrow and wide letters. Thus, in a strictly non-proportional alphabet, the *CVmedcar* value should not be subject to any distortion.

<sup>75</sup> As has already been suggested, this extended form of ‘overrun’ is a lot more pronounced than ‘simple’ overrunning, inasmuch as when the layout was created the narrow ‘double justi-

Two intersections that exert a significant influence on the frequency of word splitting are those involving the average number of characters per line (*medcar*) and the average width of the individual characters (*larcar*) (Tab. 5).<sup>76</sup>

Distribution of word splits according to the average number of characters per line and the average width of the characters				tab. 5
	Split words < 30%	Split words > 42%	Total	
<i>medcar</i>				
Up to 28 characters	48	113	161	
More than 28 characters	105	52	157	
Total	153	165	318	
<i>larcar</i>				
Up to 3 mm	62	32	94	
More than 3 mm	91	131	222	
Total	153	165	316	

**Tab. 5:** Distribution of word splits according to the average number of characters per line and the average width of the characters

When a line contains a small number of characters (fewer than 28), the manuscripts that contain the highest number of split words are a lot more common than chance would dictate, whereas the situation is almost exactly the opposite when lines are occupied by a greater number of characters. At first glance, such an effect seems to

fiction' columns constituted part of the margins, rather than part of the writing area. The behaviour of the scribes provides an idea of the ambiguous role played by these narrow columns, which ideally were conceived as empty or almost empty spaces (if intended for the placement of initials), but in practice were considered a 'transition zone' between the two opposite areas, 'black' and 'white'. A closer examination (and interpretation) of 'overrunning' and the rules that governed it would have necessitated the application of a more detailed surveying protocol that would entail making a note of the presence or absence of instances of word division in each 'overrun' line.

**76** In all three cases, the  $\chi^2$  value calculated with respect to the theoretical value is very telling: 16.5 for *larcar* (probability of 0.5%); 19.3 for *abbr* (probability of less than 1/105), and as much as 43.7 for *medcar* (probability of less than 1/107). It should be remembered that the probability threshold at 5% is 3.84.



agree with the hypothesis which states that lines that are too short increase the need to split final words. A positive correlation, albeit to a lesser degree, between the number of split words and the width of characters also emerges.

It would be an oversimplification to attribute the effect of *medcar* and *larcar* solely to the influence of a psychological factor. It should not be forgotten, in fact, that both the variables are not independent from other characteristics of the page or codex. Above all, we know that the size of characters depends to a certain extent on ruling, and therefore, indirectly, on the size of the volume in question. As regards the average number of characters per line, this is linked not only to the size of characters, but also to the length of the line, and therefore, once again, to the size of the codex. Hence it is not purely a matter of chance that when the distribution of the number of split words is cross-referenced with the size of the volume and its ruling, the value of  $\chi^2$  is always significant.<sup>77</sup> Finally, it should be remembered that all the parameters mentioned so far are strictly correlated to the density of the characters laid out on the page—in other words, the extent to which its surface is exploited.

Because of the interdependence between the different characteristics of the codex and the complex nature of each volume taken individually, it is not easy to ascertain whether or not a variable exists that is more directly linked to word division than *medcar* and *larcar*, and if so what such a variable might be.<sup>78</sup> However, we know that in addition to the size of a codex and the parameters that are contingent on the same, a non-numerical variable exists which has a much more direct impact on the length of lines, and therefore on the average number of characters accommodated by them, namely the arrangement of the text either as a full page or alternatively one laid out in two adjacent columns. The correlation between the layout of the text and the length of lines traversing it is particularly strong in Byzantine manuscripts, where a two-column arrangement was not motivated, as was the case in Western late medieval book production in the Late Middle Ages, by the desire to achieve a better exploitation of the page.<sup>79</sup> For this rea-

---

<sup>77</sup> Respectively, 8.6 (probability of 0.3%) and 4.6 (probability of 3.2%).

<sup>78</sup> The positivity of the  $\chi^2$  test which—as can be seen—provides the best result for *medcar* (see above, footnote 76), does not offer any indication in this regard, since it is limited to measuring the degree of interdependence between two variables placed in relationship to one another. This means that it is not possible to infer the type of relationship between the two variables, nor its exact nature: therefore, it is impossible to exclude the existence of a third variable which might prove to be more closely connected to the phenomenon one wishes to study.

<sup>79</sup> Concerning this issue, see Maniaci / Ornato 1995, 186–190. On the construction and utilisation of the page in Byzantine manuscripts, the author is currently working on a more wide-reaching contribution.

son, when the size of a volume is the same, a line of text in a full-page layout will on average contain at least twice the number of characters as those contained in a line of text in a two-column layout.<sup>80</sup> Therefore, if the average number of characters is so closely linked to the type of layout, it becomes necessary to verify whether or not this factor can really explain the either greater or lesser degree to which word splitting was resorted to by Byzantine scribes.

The intersection between the layout of the text and the two extreme opposites of word splitting distribution confirms this hypothesis (Tab. 6), because it reveals a very clear divergence between codices with full-page layouts—the majority of which exhibit a lower number of split words—and two-column codices—the majority of which exhibit numerous split words.<sup>81</sup>

Distribution of word splits according to text layout		tab. 6
Layout type	Split words < 30%	Split words > 42%
Full page	119	65
Two-column	35	102
Total	154	167

**Tab. 6:** Distribution of word splits according to text layout

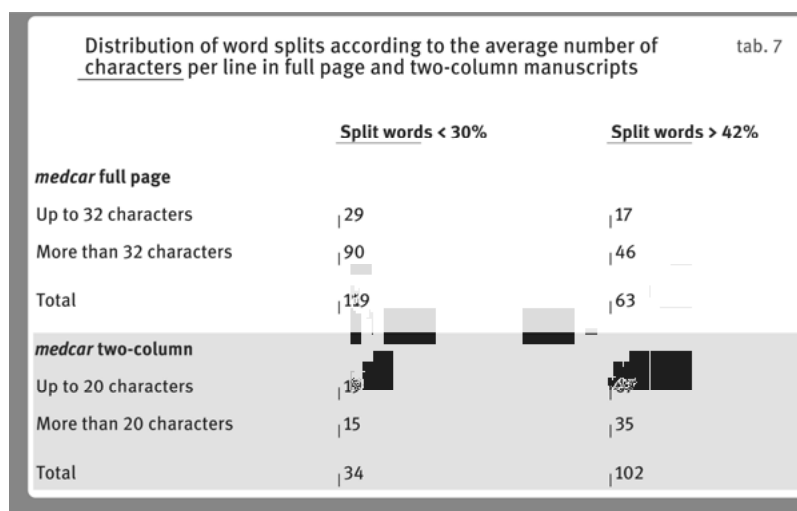
That this new variable, among those which have been examined up to this point, is the only one to exert a real influence on the distribution of split words is

<sup>80</sup> In actual fact, when the dimensions of volumes are the same, the relationship proves to be even closer, inasmuch as the sum of the width of the two columns is almost always lower than the length of a full-page line. This is ascribable to the fact that, in Byzantine codices, the construction of a two-column page, created by ‘cutting’ a vertical corridor through the justification frame, without recovering the sacrificed space—if not partially—by narrowing the two lateral margins, predominates (see Maniaci / Ornato 1995, 189). On the other hand, if one considers the corpus—which includes volumes of very different sizes—in its totality, the average relationship will be less than 2/1 in favour of full-page manuscripts. In fact, generally speaking, two-column codices are of larger dimensions.

<sup>81</sup> The  $\chi^2$  test in this case yields a result of 48.2, which represents the highest result achieved so far.

demonstrated by the fact that the effect seemingly exerted by *medcar*<sup>82</sup> is cancelled out when the sample under examination is divided into two sub-groups, based on page layout (Tab. 7).<sup>83</sup>

A similar phenomenon is produced by all the other variables, which, following an initial analysis, appeared to be directly correlated to the division of words at the end of lines (i.e. the average width of characters, volume size and ruling).



**Tab. 7:** Distribution of word splits according to the average number of characters per line in full page and two-column manuscripts

<sup>82</sup> The reader will recall that, in the case of two-column codices, the *medcar* value considered is the one that corresponds to the ‘logical line’, or in other words to only *one* column of text (at the end of which the scribe would have started a fresh one), and not the value that relates to the ‘physical line’, formed by the two flanking columns.

<sup>83</sup> The  $\chi^2$  value is 0.2 for full-page volumes and 1.1 for two-column volumes. Both values are a long way from the minimum statistical significance threshold. The discriminating properties of the ‘28 characters per line’ value—above which one notes a reduction of the number of split words—stem from the fact that in the vast majority of two-column texts the *medcar* is lower than the quoted value. In addition, as has already been stated, below this value full-page manuscripts exhibit a quantity of split words which is consistently lower than that seen in the corresponding two-column volumes. This would not be the case if the number of characters represented a decisive factor.

Conversely, all the variables that appeared to be unrelated to the phenomenon in question become operative once the full-page volumes are separated from the two-column ones. The layout of the text is therefore the key factor which, without exception, accounts for all the variables identified by the crossings carried out up till now, as the tests performed very clearly show.

When confronted with such a clear-cut—from a statistical perspective<sup>84</sup>—situation, the explanation for the relationship between the layout of the text and the number of split words is not correspondingly clear; indeed, one cannot easily imagine the tangible ways in which the scribes' conscious choices and/or spontaneous responses vis-à-vis the splitting of words might have been influenced. In other words, it is not clear whether the layout of the text should be regarded, with respect to the distribution of the split words, as a factor (of an 'automatic' or psychological nature) that has a *decisive* influence on the distribution of divided words, or alternatively simply as a *concomitant* factor, behind which is veiled the effect of one or more additional variables associated with one or the other type of text arrangement.

What then, on the material level, does the fact that a page is laid out in two columns rather than as a single block of text imply? Apart from being composed of shorter lines—which consequently contain fewer characters—a two-column arrangement entails the lines of script in the first column being delimited at their furthest point by an intercolumnar space, which is generally narrower than the outer margin that sets a boundary for the lines of script arranged in the second column. Contrary to what one might expect, neither of these two characteristics provides an explanation for the relationship that links the layout of the text to the number of split words. The hypothesis that the shortness of individual lines of script in two-column volumes automatically increases the tendency to split words at the end of lines cannot be considered valid, inasmuch as full-page manuscripts with lines of equal length—or containing an equal number of characters—present fewer split words in comparison to two-column ones. On the other hand, the hypothesis that 'overrunning' of the script into the intercolumnar space was outlawed on account of it being aesthetically displeasing—and that the 'ban' was extended, for the same reason, to the second column on the page<sup>85</sup>—would only

---

**84** This seldom happens when statistical analysis is applied in the field of historical studies, in which the phenomena—owing to the great variety of ways in which they manifest themselves, even as part of the same trend—show poorly defined profiles in the majority of cases.

**85** A survey carried out on approximately forty volumes showed that the number of split words is on average the same for each of the two columns on the page.

be confirmed if the number of ‘overruns’ were higher in full-page manuscripts, but the survey results furnish evidence to the contrary.<sup>86</sup>

What we have seen up till now permits us to exclude the possibility that the type of text layout exerted a direct influence on the attitude of copyists with respect to the problem raised by the division of words at the end of lines. On the other hand, if the text arrangement were the only factor correlated—even in a non-decisive way—to the number of split words, the separation of full-page manuscripts from two-column ones should reveal two very distinct sub-populations which are symmetrically and randomly distributed around the respective means.<sup>87</sup> The reality, however, is very different: a comparison with the random distributions shows that the adjustment is satisfactory—even if it is not perfect—only for the two-column volumes, which therefore appear to constitute a fairly homogeneous population in which a concern about splitting words at the end of lines as little as possible is practically non-existent.<sup>88</sup> In the case of full-page volumes, the situation appears to be different, because the actual distribution<sup>89</sup> highlights—with respect to the manuscripts exhibiting the greatest number of split words—a clear surplus of split words in relation to the theoretical distribution. Based on all the evidence we have at our disposal in this case, which involves two partially mixed populations, some of the full-page manuscripts follow the ‘indifferent’ trend already ascertained for two-column manuscripts, whilst in the remainder there is clear evidence of concern with respect to containing the number of split words.

## 5 Other factors that produce variations in word splitting

The situation described thus far allows us to suppose that, apart from text arrangement, various other elements can exert an influence on the distribution of

---

**86** Additionally, this hypothesis would imply a negative correlation between the number of split words and the extent of ‘overrunning’, which does not correspond to the reality observed in our sample (see above, 627–628).

**87** In agreement with the corresponding mean binominal distributions. The scattering around the averages in this case is due to the slightly different choices made on the part of the scribes and the inherent fluctuations within the sample (the survey was limited to 45 lines of text).

**88** The average for the two-column codices is, in fact, 41.8%, and therefore very close to the random value (42.1%).

**89** On average, 35.1%.

split words. Since we have already excluded the impact of qualitative and graphic variables, it only remains for us to consider data—albeit only partial, but in any event numerically sufficient—relating to abbreviations.

In effect, abbreviations prove to be negatively correlated to the number of split words, inasmuch as the scribes who copied manuscripts that exhibit many abbreviations did not split words at the end of lines very often (Tab. 8).<sup>90</sup>

Distribution of word splits according to the abbreviation rate			tab. 8
Abbreviations (386 mss)	Split words < 30%	Split words > 42%	
Up to 7 abbreviations (2.3%)	42	73	
More than abbreviations (2.3%)	44	19	
Total	86	92	

**Tab. 8:** Distribution of word splits according to the abbreviation rate

It will be useful, then, to expand our knowledge of the phenomenon. This can be accomplished by confirming the possible existence of a direct link between the layout of the text and the abbreviation rate, and by attempting to clarify, should one exist, whether such a link—as we have seen before with the other variables—is entirely responsible for the correlation between the abbreviation rate and the number of split words. By crossing data relating to the text arrangement and the abbreviation rate it can be seen that the distribution of the two variables is not independent: the codices that present with an abbreviation rate above 2.3%<sup>91</sup> exhibit, in fact, full-page layouts in a significant majority<sup>92</sup> (Tab. 9).

<sup>90</sup> In the table, the  $\chi^2$  value for the abbreviations should not be related to the figure previously cited for the average number of characters per line (see Tab. 7), because the value for this variable is a lot lower and limited to only two of the centuries considered. What counts more than anything is the probability value, which, even in the case of abbreviations, ensures that we are not dealing with random variations.

<sup>91</sup> This figure, chosen on an experimental basis as a dividing line, corresponds to the number of abbreviations occurring in 60 words of text, equivalent to approximately 300 characters. The absolute value is 7 abbreviations (see Tab. 8–10).

<sup>92</sup> The  $\chi^2$  value is 13.9 (probability = 0.02%). Tab. 9 also reveals the fact that, in the 11<sup>th</sup> and 12<sup>th</sup> centuries, abbreviation-rich Byzantine manuscripts are clearly in the minority.

Abbreviation rate according to page layout			
Percentage of abbreviations	Full page	Two-column	Total
< 2.3%	117	148	265
> 2.3%	78	43	121
Total	195	191	386

tab. 9

**Tab. 9:** Abbreviation rate according to page layout

Therefore, it can be stated that there exists in the Byzantine manuscript—at least in the 11<sup>th</sup> and 12<sup>th</sup> centuries—an association between the full-page layout and an abundance of abbreviations.<sup>93</sup> However, this association is not sufficiently strong to ascribe the impact of the rate of abbreviations on the number of split words to a simple ‘structural effect’.<sup>94</sup> If one analyses the behaviour of codices exhibiting ‘slight’ and ‘considerable’ rates of abbreviation in relation to text arrangement (Tab. 10), the following facts emerge:

1. The abbreviations exert an effect on the degree of word splitting irrespective of the type of text arrangement: the group composed of ‘two-column codices containing many abbreviations’ is, in fact, accompanied by a reduction in the number of abbreviations, whilst the opposite group composed of ‘full-page codices containing few abbreviations’ registers a slight increase in relation to the overall average for full-page volumes.
2. The effect of the data partitions, whose two components act in the same way, is increased: the two groups composed of ‘full-page codices containing many abbreviations’ and ‘two-column codices containing few abbreviations’ produce percentages of split words which are lower and higher, respectively.
3. The effect of the type of text arrangement is in any event more relevant than that of the rate of abbreviations. Indeed, it can readily be seen that the re-

<sup>93</sup> This is an association which, needless to say, should be viewed as an overall trend. This does not discount the fact that 36% of the volumes with an abbreviation rate exceeding 2.3% have two-column layouts.

<sup>94</sup> The ‘structural effect’ should be regarded as an only seemingly significant result—and difficult to ‘mask’ as such—generated by the sample’s characteristics, which mean that the link between two variables, which outwardly appears to be direct, in reality does not exist, or alternatively is induced by a third variable not identified as being responsible. See Maniaci / Ornato 1993, 22.

sult obtained when one cross-references ‘few abbreviations/full-page’ (divergent trends) is lower than the overall average of split words seen across the entire corpus. The same observation is true—*mutatis mutandis*—when one cross-references ‘many abbreviations/two-columns’,<sup>95</sup> whose average value is higher than the overall one.

Distribution of word splits according to layout and abbreviation rate			
Percentage of abbreviations	Full page	Two-column	Total
< 2.3%	35.6%	42.3%	39.4%
> 2.3%	31.7%	38.4%	34.1%
Total	34.1%	41.5%	37.7%

**Tab. 10:** Distribution of word splits according to layout and abbreviation rate

The fact that a cause and effect relationship does not exist—in either sense—between text arrangement and abbreviations, supports the hypothesis that their interaction with the phenomenon of word splitting at the end of lines can be ascribed to the concomitant action of the three parameters that characterise Byzantine book production: the scribes who exercised deliberate control over the division of words favoured the simplest type of page layout and their reading skills—and/or those of their commissioning clients—were compatible with a certain number of abbreviations. It is only natural to suppose, then, that one or more factors unrelated to the codex’s material aspect had something to do with this preference, namely historical period, cultural context and readership. Unfortunately, our level of knowledge concerning the dissemination of written culture in the Byzantine world—and in particular knowledge with respect to localisation—prevents us

<sup>95</sup> A detailed examination of the distribution of split words—carried out with help of a diagram not reproduced in the present contribution—makes it possible to gain a very clear understanding of the relationship that links this phenomenon to abbreviations. In full-page layouts, in fact, one observes that the bulk of codices in which the percentage of split words is lower than 30% simultaneously present a higher number of abbreviations. Conversely, in two-column layouts one notes that the manuscripts exhibiting a rate of abbreviations in excess of 2.3% belong to two sub-groups which can be distinguished without ambiguity based on the number of split words they contain, depending on the care taken or indifference shown towards controlling this particular aspect of line management.



from being able to divide the samples into sub-groups which are at one and the same time well defined and sufficiently large so as to make it possible to subject the hypothesis to a statistical verification. In practice, the only data fields that can to a certain extent be made use of are those which relate to the manuscripts' century of manufacture and their textual typology.<sup>96</sup>

The distribution of split words in relation to the most representative textual typologies within the corpus shows that the content of volumes effectively constitutes a discriminating factor (Tab. 11).

Text types	% Split words	Split words < 30%	Split words > 30%
Bible	33.1%	21	9
Biblical	38.3%	17	22
Hagiography	42.9%	4	16
Homiletical	41.5%	18	39
Liturgical	37.7%	31	31
Patristic	37.2%	19	19
Secular	33.3%	29	15
Total	37.8%	139	151

**Tab. 11:** Distribution of word splits according to text type

Two text typologies—Bibles and secular codices—stand out from all the others on account of the scarcity of split words they exhibit, whereas hagiographies and homilies stand out for the opposite reason. It is therefore logical, at this point, to pose the question as to whether or not, and how content might interact with the

<sup>96</sup> Concerning the systematic classification of the codices based on their content, the recent work by Sautel 1995 was referred to, in particular pages 12 to 13, where various sub-groups are consolidated so as to obtain groups of sufficient size (for the complete list of the classes adopted, see Tab. 11). Despite the fact that the inventory edited by Sautel presents the shortcomings of any database that covers a large span of time and is enhanced with information of heterogeneous origins, it still represents a valuable source of information relating to almost 4,000 Greek manuscripts, many of which have not yet been scientifically catalogued. In this regard, see the reflections formulated in Maniaci 1996, the response by Sautel 1996, and the observations made by Muzerelle / Ornato 1997, 26–36.

two parameters—text layout and abbreviations—that we already know are linked to the number of split words.

The cross between content, text arrangement and the number of abbreviations (Tab. 12) produces similar results to those obtained for the interaction between layout and abbreviation rate: the effect of the textual type on the number of split words does not decrease simply due to its more or less large representation, within one or another type, of a given text arrangement or a particular abbreviation rate.

In fact, within each sub-group, the Bibles and secular texts exhibit, on average, fewer split words in comparison to codices of other textual content, whereas hagiographies and homilies contain a greater number.

Distribution of word splits according to text type, layout and abbreviation rate				tab. 12
Text types	Full page	Two-column	< 2.3% abbr.	> 2.3% abbr.
Bible	28.3	44.9	33.7	31.5
Biblical	34.7	41.6	39.9	36.2
Hagiography	40.6	43.5	43.4	42.2
Homiletical	38.0	42.9	42.6	36.8
Liturgical	33.6	40.4	37.9	37.4
Patristic	37.4	36.6	40.7	32.6
Secular	32.1	37.8	36.4	32.0
Total	34.1	41.5	39.3	35.2

**Tab. 12:** Distribution of word splits according to text type, layout and abbreviation rate

However, within each textual typology—if one excludes hagiographic codices, which appear to totally ignore the problem of word splitting<sup>97</sup>—the text layout and abbreviation rate maintain, albeit in varying degrees, their influence.

<sup>97</sup> In order to understand the reasons at the root of this anomaly it would be necessary to carry out a closer analysis of the group in question, behind whose apparent unity of content—a product of the necessity to adopt, in order to carry out a preliminary examination of the phenomenon in question, very broad classes identified by highly generic ‘labels’—are concealed, in all likelihood, profound typological differences.

The sub-division of the manuscripts according to the century of their production (Tab. 13) also exposes some differences vis-à-vis the frequency of word splitting phenomenon.

Distribution of word splits according to centuries		tab. 13
Century	Split words < 30%	Split words > 42%
9 <sup>th</sup>	10	3
10 <sup>th</sup>	26	53
11 <sup>th</sup>	45	78
12 <sup>th</sup>	54	28
13 <sup>th</sup>	13	5
14 <sup>th</sup>	5	0
Total	153	167

**Tab. 13:** Distribution of word splits according to centuries

In particular, codices of the 9<sup>th</sup> century on the one hand, and those of the 12<sup>th</sup> to 14<sup>th</sup> centuries on the other, exhibit fewer split words compared with the intervening centuries. In this case, too, by crossing the chronological factor with previously<sup>98</sup> considered factors (Tab. 14), the differences between centuries do not disap-

**98** Given that certain centuries (the 9<sup>th</sup>, 13<sup>th</sup> and 14<sup>th</sup>) are rather sparsely represented in the corpus, the comparison was carried out by grouping together the centuries into just two classes, based on behaviour already identified during the comparison of split words. The same procedure was adopted with respect to textual typologies (Bibles and secular texts on the one hand; hagiographies and homilies on the other). It goes without saying that the said combinations were devised purely for the sake of convenience, and that they neither point towards nor influence an explanation for the phenomenon that interests us here. In this connection, it should be noted that a fundamental difference exists between the sub-divisions created on the basis of page layout and abbreviation rate, and the groupings that concern the dating and content of the codices. In the first case one is dealing with 'natural' populations endowed with their own significance and created on the basis of defined characteristics which can prove to be more or less correlated to the phenomenon—in our case, word splitting—one is seeking to explain. By contrast, in the second case one is dealing with 'artificial' and, as it were, 'tautological' agglomerations, inasmuch as they are created using as a starting point the very phenomenon that the analysis is focused on.

pear, even if, at the same time, the tendencies already identified with respect to text arrangement, content and abbreviation rate remain the same.

Century	Full page	Two-column	< 2.3% abbr.	> 2.3% abbr.	Bib + Sec	Hag + Hom
10 <sup>th</sup> –11 <sup>th</sup>	36.84	41.82	40.48	36.76	35.24	41.37
9 <sup>th</sup> ; 13 <sup>th</sup> –14 <sup>th</sup>	31.86	38.40	36.79	31.33	31.61	37.97

**Tab. 14:** Distribution of word splits according to centuries, layout, abbreviation rate and text type

If one considers simultaneously all the variables that minimise or maximise, respectively, the number of split words, it will be found that they are often concurrently present in the one and the same manuscript.<sup>99</sup> Conversely, an association between two or more characteristics which act in the same way has the capacity to ‘enhance’—either positively or negatively—the effect exerted by each characteristic individually.<sup>100</sup> Thus, the coexistence of any three of the four sub-groups that minimise word splitting (full-page codices/codices containing many abbreviations/codices containing biblical or secular texts/codices dating from the 9<sup>th</sup>, 12<sup>th</sup> and 13<sup>th</sup> to 14<sup>th</sup> centuries) results in a reduction of the average percentage of split words to around 29% (and the coexistence of all four criteria lowers it even further to 25%).<sup>101</sup>

**99** Some examples: our corpus contains 52 volumes that share the following characteristics: (a) all were transcribed in either the 10<sup>th</sup> or 11<sup>th</sup> centuries; (b) all have hagiographic or homiliary content; (c) all have two-column layouts; and (d) all have an abbreviation rate lower than 2,3%. On the other hand, the volumes that—leaving aside three of the characteristics listed above—have an abbreviation rate of above 2.3% number only 7, and those produced in any one of the other four centuries number 5, as do those of biblical and secular content. Finally, 11 have full-page layouts.

**100** For example, the group of codices with full-page layouts, with an abbreviation rate above 2.3%, and which were transcribed in the 9<sup>th</sup>, 12<sup>th</sup>, 13<sup>th</sup> and 14<sup>th</sup> centuries (a total of 45 volumes), is composed entirely of manuscripts whose scribes clearly took trouble to limit the division of words.

**101** It is true, however, that in this case the number of volumes descends to just 12. The synergy produced by the association of multiple factors is considerably less evident in the sub-groups that generated the higher frequency of split words, inasmuch as there is an upper limit that corresponds to an absolute lack of concern regarding the problem on the part of the scribes.

## 6 A few interpretations worth considering

The close interaction between the parameters considered up till now makes it difficult to access on an individual basis the effect that each of them exerted on the management of word splitting. In fact, it is not easy to isolate the effect exerted by an individual factor, inasmuch it will normally present in association with others which act in the same way or in the opposite way, thereby contributing, in varying degrees, to the intensification or diminution of the particular effect. Indeed, among full-page volumes, many can exhibit an abbreviation rate of more than 2.3%; a certain number of such volumes will also be of biblical or secular content, and so on.

In order to determine the importance of the individual factors, two different procedures can be applied, each of which has advantages and disadvantages.<sup>102</sup> Fortunately, the application of one or the other of the two procedures produces the same result: the element that has the greatest influence on word splitting turns out to be page layout,<sup>103</sup> followed by textual content, whereas the dating and number of abbreviations appear to be less important, despite being partly independent from the other two factors. The dating and abbreviation rate exert their respective effects above all on full-page volumes. This finding is hardly surprising if one considers that two-column manuscripts constitute—as we have seen—a more homogeneous sub-group and one which tends to show indifference to limiting the division of words.<sup>104</sup>

---

**102** The first criterion—which makes it possible to isolate the effect of a given parameter—consists in limiting the calculation of the percentage of split words solely to the volumes in which all the other parameters act in a contrary way (for example, codices with full-page layouts, but which contain hagiographic material or present an abbreviation rate of less than 2.3%, and which were transcribed in the 10<sup>th</sup> or 11<sup>th</sup> centuries). The second method consists in counterposing two at a time the variables that act in the opposite way (full-page/few abbreviations, two-column/biblical or secular content, etc.), and observing whether the average percentage of split words in relation to the overall average of the corpus increases or decreases. The most active variable will obviously be the one which most frequently deviates the result of the comparison in the direction that complies with its autonomous tendency. The second method does not enable one to isolate on a case-by-case basis the effect of an individual variable, whose impact therefore has to be evaluated based on the outcome of all the comparisons.

**103** This is true for all the comparisons, with the exception of one—full-page manuscripts and hagiographic/homilary content—in which the latter factor is dominant (with a very small variation: 37.85% as opposed to 37.58%).

**104** Whilst the desire to limit word splitting can manifest itself in varying degrees and can even extend to eliminating them entirely (or almost entirely), the lack of any concern for the

Even if the variables examined up to this point do not contribute to the definition of a sufficiently characterised and homogeneous ‘codex type’—based on structure, content and quality of craftsmanship—their convergent trends make it possible to hypothesise that the need to limit the division of words at the end of lines was closely linked to the existence of specific commissioning and consumption environments.

In particular, it is only natural to suppose that a concern to limit the number of split words was felt mostly by readers who paid closer attention to the correctness and legibility of the text, rather than to the sumptuousness of its presentation. This attitude undoubtedly characterised books created for study purposes or for personal reading by and/or for a cultivated readership. The predominance of secular texts, the preference for a simpler layout (i.e. full-page), the tendency towards an intensive use of abbreviations, a preference for cursive or tendentially cursive (*‘corsiveggiante’*) scripts and/or less standardised ones, rather than for so-called ‘Perlschrift’ and ‘traditional’ scripts,<sup>105</sup> are all characteristics that appear to support this hypothesis. In other words, the manuscripts produced for private consumption by a readership of a higher cultural calibre exhibit less concern for their appearance—i.e. the quality of the writing support and the aesthetic principles used for the page’s layout and its decoration—and a greater concern with respect to their substance, which is to say the integrity and legibility of the text. This could explain the almost total absence of a correlation between the number of split words and the richness of decoration: in fact, limiting the number of di-

---

phenomenon cannot be progressive and produces, *ipso facto*, a concentration of values around the level that corresponds to the phenomenon’s random modes. Concerning the effect of the different variables, the page’s proportion should be added to the list: values lower than 0.75 correspond to a lower number of split words (36.2%), whereas values above 0.75 register a higher rate (38.5%). Clearly, the gap is too small to make it worthwhile to attempt an explanation. In any event, this is a systematic effect which proves to be independent of other factors, and in particular from the layout of the text, whose influence on the proportion of the page is in any case smaller than in the Latin context (see Maniaci / Ornato 1995), and from the dating of the codices (although the proportion tends to become narrower over time: see Maniaci 1995, 31 footnote 35). It should be pointed out that the effect—which for now remains inexplicable—apparently exerted by page proportion only manifests itself in the group that registers a low percentage of split words.

**105** The term, coined by Herbert Hunger in the mid-1950s (Hunger 1954), is adopted here, in a very broad sense, to indicate the ideal calligraphic model which prevailed from the middle of the 10<sup>th</sup> century until roughly the middle of the 12<sup>th</sup>. Old-fashioned scribes resisted this script, especially for reasons of provincialism and/or lack of skill, in a show of loyalty/deference to earlier epochs; see Canart / Perria 1991, 83–87.

vided words became the chief priority on the list of requirements demanded by readers more concerned about content than the appearance of their books.

However, there is a specific case that seems to contradict this interpretation. One of the textual typologies whose layout (largely speaking) is highly meticulous and richly decorated—the Bible, mostly represented in the corpus by the New Testament—contains the lowest number of split words (33.6%). The similarity between biblical texts and secular texts—both of which testify to a rather marked tendency towards limiting word division—is not matched by other features of their manufacture; indeed, the two groups, far from constituting a homogeneous whole, tend on the contrary to run counter to each other with respect to their palaeographic and codicological characteristics (Tab. 15).<sup>106</sup>

Amongst other things, the manuscripts of biblical content are superior in the quality of their writing support and their pages are less exploited. They are also more uniform and contain far fewer abbreviations. With respect to certain parameters (i.e. parchment quality, ruling complexity and spacing, and percentage of ‘blackness’), they exhibit the best values bar none among all the other groups. Finally, one notes that the behaviour of the Bibles in relation to split words is anomalous in comparison to that of the other textual typologies, in the sense that it remains almost constant in all witnesses of all periods, whether they be slightly or greatly abbreviated (Tab. 16).<sup>107</sup>

---

**106** A few of the variables entered in the table require some clarification as regards the values they have conventionally been measured by. The ‘quality of the parchment’ is evaluated on the basis of a hierarchical scale as follows: 2 = excellent; 1 = good; -1 = mediocre; -2 = poor. The ‘ruling complexity’ index equates to the sum of all the supplementary lines (i.e. those which are not indispensable to the delimitation of the writing area), with the exception of the narrow columns used for double justification. The ‘exploitation of the parchment’ indicates the mm<sup>2</sup> ideally available for each character; these are calculated by dividing the total surface area of the page (and not only that of the writing area) by the average number of characters it contains. The ‘normalised ruling unit’ expresses the relationship between the ruling and a volume’s size (multiplied by 10,000). The ‘irregularity of the script’ is afforded a rough evaluation by calculating the sum of the coefficients of variation of the height and width of the Greek letter *omicron*. The degree of ‘overrun’ was considered for only 228 codices endowed with two-column layouts, taking into account only the percentage of cases where the script exceeds the outermost vertical of the framing column. Finally, the ‘rate of abbreviation’ parameter—relating, as the reader will recall, to only 386 manuscripts—is calculated, as previously stated, based on a sample composed of 60 words.

**107** Two-column volumes represent an exception, however, since they are largely in the minority (5 in 63, or 8%).

Differences between the Bible and secular texts with respect to palaeographic and codicological characteristics			
Text types	Bible	Secular texts	Entire corpus
Parchment quality	0.42	-0.09	0.03
Ruling complexity	3.59	2.16	2.68
Percentage of 'blackness'	29.0	31.0	30.1
Parchment utilisation	56.62	46.6	61.74
Normalised RU	16.11	14.21	14.59
Writing irregularity	19.67	23.66	20.86
Abbreviations	5.53	18.53	7.93
Overrun	9.86	5.78	8.32

tab. 15

**Tab. 15:** Differences between the Bible and secular texts with respect to palaeographic and codicological characteristics

Distribution of word splits in bibles and other text types according to layout, abbreviation rate and centuries		
	Bible	Other texts
Full page / two-column	30.3	42.9
< 2,3% – > 2,3% abbr.	31.1	33.8
Centuries 10 <sup>th</sup> – 11 <sup>th</sup> / 9 <sup>th</sup> ...12 <sup>th</sup>	34.1	32.6

tab. 16

**Tab. 16:** Distribution of word splits in bibles and other text types according to layout, abbreviation rate and centuries

The apparent anomaly that the manuscripts with biblical content show in the management of word division could be explained by the particularly high quality of their preparation; indeed, in a context where every last detail of the volume's manufacture had the utmost attention paid to it, even a relatively secondary concern such as word division would have been carefully considered.



## 7 Treatment of divided words in unusual situations: at the end of pages and at the end of quires

If we accept that the attitude of copyists towards the splitting of words springs not from two different situations—i.e. an awareness or ignorance of the role it plays in legibility—represented to a greater or lesser extent in the volumes examined, but instead from a complex interrelationship between the different requirements that contribute to the ‘planning’ of each individual manuscript, one also has to consider that in one and the same codex the word splitting phenomenon can manifest itself in different ways, depending on the particular circumstances involved.

The problems that result from the steady application of a ‘re-assemblage’ procedure for the reading of a text arranged in a series of lines positioned one on top of the other have already been touched on in the introductory remarks. Needless to say, the longer and/or more irregular the visual pathway that the eye has to follow is, the more troublesome the procedure becomes. This is certainly the case with the final line on a page which, representing a particularly sensitive juncture in the reading flow, and therefore one which can be a source of potential errors, naturally calls for greater attention to be paid to it with respect to word splitting. Furthermore, because final lines are far less numerous than all the other lines which form a complete text, without doubt less time and effort has to be expended on them.

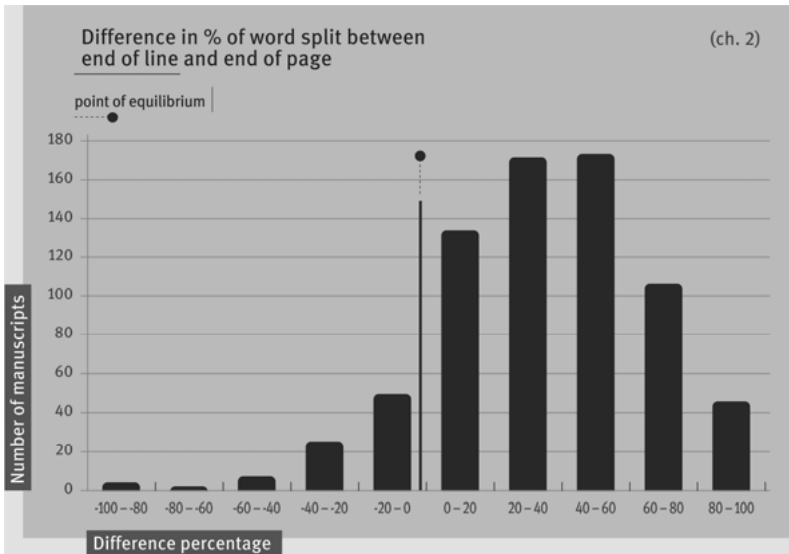
If one calculates separately the percentage of split words located at the end of pages, one finds that in actual fact their distribution is rather different from the distribution of those located at the end of lines, in that it is weighted towards considerably lower values (Chart 2). The same observation is valid not only for all the manuscripts taken as a whole, but also if one calculates, for each individual codex, the difference between the two parameters.<sup>108</sup> This means that manuscripts in which the scribe shows himself to have been unconcerned about split words at the end of pages, but on the other hand sought to avoid those at the end of lines, do not exist. Moreover, the scribe also shows that he was alert to the phenomenon in general and tended to pay more attention to the special situation represented by the end of a page (Tab. 17).

---

**108** In only 80 cases out of a total of 700, or 11.4%, was a larger number of split words at the end of pages than at the end of lines recorded. However, in these cases, too, the difference never exceeds 20%, and proves compatible with the random variation. In addition, the manuscripts in question share no characteristics in common, which would explain their concordance in the handling of word division.

Distribution of word splits at the end of the page according to their distribution at the end of lines					tab. 17
% split words, end of line	% split words, end of page				Total
	< 20%	20%–30%	30%–42%	> 42%	
< 20%	137	18	10	10	145
20%–42%	159	111	166	126	362
> 42%	69	101	81	42	293
Total	265	220	147	68	700

**Tab. 17:** Distribution of word splits at the end of the page according to their distribution at the end of lines



**Chart 2:** Difference in % of word split between end of line and end of page

While almost half of the scribes (293, or 42%) were totally unconcerned about the number of split words occurring at the end of lines, only about a tenth of them

(9.7%)<sup>109</sup> adopted the same attitude towards split words occurring at the end of pages. Needless to say, the scribes who were careful to avoid splitting words at the end of lines were even more careful to avoid doing so at the end of pages. However, also among those who did not hesitate to split words at the end of lines there was a clear reluctance to do likewise at the end of pages.<sup>110</sup> Hence there existed a significant number of scribes who were concerned about limiting split words only when it came to transcribing the final line of a page.<sup>111</sup>

These observations highlight the fact that in Byzantine books transcribed in minuscule script, already starting in the 9<sup>th</sup> century split words were perceived as an occurrence that could interfere in a negative way with the legibility of the text; and that the attitude of the scribes with respect to remedying the problem was commensurate with, at one and the same time, the degree of potential detriment to the text and the ‘expenditure of energy’ required to carry out the remedy. Indeed, more than 90% of the scribes paid a certain amount of attention to the phenomenon, but a little under 60% of them did it at each line’s end,<sup>112</sup> albeit a lot less keenly. In other words, a codified norm existed that proscribed the splitting of words at the end of a page (it should be noted that only 42 scribes, equating to 6%, completely neglected to observe this rule) which extended, as far as possible, to the limitation of split words at the end of lines, at the discretion of the scribe.<sup>113</sup>

---

**109** Or, more correctly stated, 293 codices. The equivalence is, however, legitimate, because during the surveying phase for each codex an effort was made to carry out all the calculations on sheets or quires that were in all likelihood executed by the same hand.

**110** Strictly speaking, one has to deduce that only 6% of the scribes (42 in a total of 700) appear to be entirely resistant to maintaining any voluntary check whatsoever over split words. The reality is, however, somewhat gradated, inasmuch as the result of the survey (carried out, we should recall, on 45 lines in the case of split words at the end of lines and on three quaternions for those located at the end of pages) is influenced to a certain extent by random fluctuations, and by the fact that the threshold set at around 42% cannot have an absolute value. However, since the fluctuations occur in both directions, the percentage of ‘unconcerned’ scribes will not be all that far off the value indicated.

**111** It is interesting to observe that this concern was acutely felt, given that in 58% of the manuscripts where the split words reach random values the number of split words at the end of pages falls to under 30%, and in 24% of cases to under 20%.

**112** A more in-depth investigation—based on a closer examination of the individual codices in which an effort to limit split words proves to be obvious—could make it possible to identify fluctuations in the amount of attention paid by the scribes to the issue, and possibly to link such variations to ‘cyclical’ factors (first quire/other quires; first leaf of a quire/successive leaves; first side of a leaf/second leaf; first lines of a page/last lines, etc.).

**113** The analysis of the distribution of split words at the end of pages in relation to the other external and internal characteristics of the codex reveals that the attention paid by the scribes is not uniform in all the sub-groups, but instead is distributed proportionately in relation to the

The grouping together of the scribes who did not apply—and who were probably not aware of—any rule aimed at regulating the division of words is not the product of pure chance. Indeed, within this sub-group can be found almost all the oppositional factors linked to the characteristics of the codex that have already been shown to exert an effect on the number of split words. However, one of these elements seems to be particularly pronounced, namely the rate of abbreviations, which is much lower in the manuscripts that exhibit a total lack of concern on the part of the scribes vis-à-vis the splitting of words.<sup>114</sup> In addition, the scribes who showed no concern whatsoever towards splitting words also employed parchment of lower quality and showed a clear preference for two-column layouts. All these elements, even if they cannot be arranged into an unambiguous whole, do not identify a specific type of codex, but rather a level of production which is noticeably sloppy, and one which is certainly a far cry from cultivated attitudes and concerns.<sup>115</sup>

Up to this point, the divided words located at the end of pages have been analysed as if they were all the same. In reality, this is not the case, since the alternation of *recto* and *verso* sides in a manuscript results in two different situations arising. In one situation both the scribe and the reader are obliged to turn the page in order to continue, respectively, either the transcription or the reading of the text. In the other setting, whereas the reader's eyes can shift naturally and without interruption to the neighbouring page, the scribe—if we take for granted that the transcription unfolds in a natural sequence on pre-cut bifolia<sup>116</sup>—has to move from one bifolium to the

---

total percentage of split words occurring at the end of lines, according to the pattern already outlined.

**114** Only 6.9% of the codices displaying a 'total lack of concern' (in fact, only three items) have a rate of abbreviation higher than 2.3%, as opposed to 31.3% across the entire corpus.

**115** As has already been observed, among the manuscripts containing secular material, those with full-page layouts and many abbreviations predominate. Not by chance, this typology—which is associated with a learned readership—is heavily underrepresented in the group of volumes which exhibit a lack of concern towards the splitting of words.

**116** For now, we need not pose the question as to whether or not the transcription was carried out on separate bifolia, or instead on an already folded and assembled quire. We can exclude, at the outset, the possibility that the copying took place in a natural sequence, but was instead carried out by means of a procedure known as 'imposition', because this—also in the case of the manuscripts in Latin script which provide the only examples—represents a late and minority practice that was pointed out for the first time in 1928 by Charles Samaran (see Samaran 1928), upon which a vast bibliography has accumulated, consisting, in most cases, of reports of new examples which have come to light purely by chance. The main references are listed in Ruiz Garcia 1988, 171, and Lemaire 1989, 214, to which we can add important contributions by Gilissen 1977, 114–122 (with a bibliography on the subject on pp. 117–118), and by Bozzolo / Ornato 1980, 154–175. No indisputable cases of 'imposed' Greek manuscripts are known to exist; for an analysis

next. The second possibility becomes even more apparent when a change of bifolium coincides with a change of quire.

In this scenario it is reasonable to suppose that the occurrence of a split word at the end of a page will result in more serious consequences if it is accompanied by a transition from the *recto* to the *verso* of the same leaf. Indeed, in this case the reader is constrained to memorise, at the moment he or/she turns the page, the first fragment of the split word, with the risk that the ‘re-assemblage’ procedure will be disrupted. In the opposite case, where the reader has instead before his/her eyes both halves of the split word, the ‘re-assemblage’ procedure, despite being less easy because it occurs on the same page, will in any event result in less disruption.

If the hypothesis coincides with reality, one should expect to observe an imbalance between the number of *recto/verso* word splits and *verso/recto* ones, in favour of the latter. Verification of this postulate is very simple to carry out at a global level: it is sufficient to make separate counts of the number of split words that occur in either one or the other position.<sup>117</sup> The result obtained, with a significant variation of 17.4% in favour of *verso/recto* caesurae (4,414, as opposed to 3,761), fully supports the hypothesis.<sup>118</sup>

By contrast, it is impossible to verify the validity of the same hypothesis for every manuscript, since the number of split words occurring at the end of pages—as has already been ascertained—is, as a rule, insufficient to render it viable to subject it to a statistical analysis. However, one can attempt to establish—as in the case of split words at the end of lines and at the end of pages—whether or not the phenomenon presents in variable ways in relation to the characteristics of the codices. The outcome, in this instance, is contrary to what one might expect: indeed, while the quantitative imbalance between *recto/verso* and *verso/recto* split words persists—thereby confirming the generalised dissemination of the phenomenon—its extent, between one group and the other, does not present variations worthy of comment (in contrast to what emerged vis-à-vis other aspects of the distribution of split words). Furthermore, its extent is independent from the total average number of split words. Therefore, the doubt arises that this particular phenomenon does not in fact derive from the more or less widespread application of a rule intended to aid reading, but rather from a reaction on the part of the copyist, who would intentionally tend to limit errors which could potentially be fostered by the fact that, when

---

(with a negative conclusion) of two questionable cases of small paper codices dating from different periods (the 12<sup>th</sup> and 15<sup>th</sup> centuries), see the short piece by Irigoien 1992.

**117** The survey was carried out on three quires (all quaternions) of each codex, located at the beginning, middle and end of the gatherings.

**118** The  $\chi^2$  test yields a result of 52, which is a very high value.

turning a page, visual control on the previous line is lost. It has already been stated that such a reaction is intentional rather than unwitting, since one observes that subscribed letters are also—in the codices where they are present—consistently more numerous between the *recto* and *verso* than in the opposite situation.<sup>119</sup> Given that the practice of subscribing the final letters of a word ('codino') represents an 'extreme' solution deployed in order to avoid word splitting—when it is impossible to do otherwise—it is clear that its quantitatively differentiated use on the *recto* and *verso* and *verso* and *recto* of sheets reveals the existence of an outright rule, aimed perhaps—upon first consideration, at least—at avoiding transcription errors rather than promoting legibility.

One could object that, from the scribe's point of view, there should be no great difference between the two types of page change, since the transition from a *verso* to the following *recto* is also determined by a continuity solution, owing to the shift from one bifolium to another. However, this objection does not hold up if one supposes that the scribe wrote not on loose and separate bifolia, but instead on an already folded and assembled quire, even if it was not necessarily already sewn.<sup>120</sup>

Finally, the succession of *verso/recto* alternations can be upset by another continuity solution: that which involves juxtaposing the last *verso* of a quire with the first *recto* of the successive one. If one looks closely, in contrast to the normal transition from *verso* to *recto*, the word split that happens to fall in this position constitutes a rather special event, and one which is laden with ambiguity from the outset. On the one hand, one can imagine that the scribe regarded the opening<sup>121</sup> between the end of one quire and the beginning of the next in just the same way that he regarded those occurring within one and the same quire, and therefore entirely like a normal instance of word splitting. On the other hand, one can also suppose that, on the contrary, the end of a quire was perceived as a strong break, and that each quire

---

**119** The overall average (calculated from three quires) is, respectively, 4.62 for *recto/verso* 'codini' and 3.08 for *verso/recto* ones. The  $\chi^2$  test on the overall averages again yields a result of 52, which is highly significant. This phenomenon is also present in all the categories into which the corpus was sub-divided.

**120** The hypothesis could be further supported if the number of split words at the end of lines that coincide with the inner side of the central bifolium turned out to be the same as that seen in all the other cases. In order to verify this, it would be necessary to make a separate count of the split words that fall in this position.

**121** The term 'opening' corresponds to the French 'double page', according to the definition provided by Muzerelle 1985, 92; see Maniaci 1996 (1998<sup>2</sup>), 126.

constituted, in some respects, an independent unit, even when it did not contain a distinct section of a text.<sup>122</sup>

The second option is the one that was settled on. A count of all the instances of split words falling at the end of quires reveals that they are proportionately less numerous not only than the remaining *verso/recto* splits, but also the *recto/verso* ones.<sup>123</sup> That said, the phenomenon's trend is different from the one observed vis-

---

**122** The scribes of the Latin Evangeliaries that were produced between the 7<sup>th</sup> and 11<sup>th</sup> centuries show themselves to have been abundantly aware of the need to match the material structure of the codex with the canonic partitions located between one Gospel and another, and between the Gospels and the material that normally accompanies them (i.e. Prologues, Eusebian Canons, summaries, *capitula* or Breviary, *Capitulare Evangeliorum*). This demonstrates the very widespread habit of starting the transcription of each individual textual unit—and, in particular, of each of the four Gospels—on a fresh quire, and concluding, if necessary, the previous text (possibly including the *capitula* of the successive Gospel) on a quire whose thickness is different from normal): Bischoff 1994. I have identified an adherence to the same practice in the various Greek Evangeliaries dating from the 9<sup>th</sup> to 12<sup>th</sup> centuries. Here, I cite, purely as examples, Vat. Pal. gr. 220, 9<sup>th</sup>–10<sup>th</sup> centuries; Vat. gr. 1159, 10<sup>th</sup> century; Ott. gr. 297, 10<sup>th</sup>–11<sup>th</sup> centuries; Pal. gr. 227, 12<sup>th</sup> century, and Roma, Biblioteca Vallicelliana, B 133, 12<sup>th</sup> century (in these codices the portraits of the evangelists, when present, are painted as miniatures on the *verso* of the last page of the quire where a Gospel ends, or on the *verso* of an independent bifolium that holds the *capitula*, or alternatively on the *verso* of a loose leaf with a stub. In contrast, the more recent Vat. gr. 1160 (13<sup>th</sup>–14<sup>th</sup> century) does not present a continuity solution between one Gospel and another (the portraits of the evangelists are also executed on the *verso* of a page—not necessarily the final one—forming part of a quire of text). We shall not concern ourselves here with establishing whether the perception of a quire as a complete and independent unit corresponded solely to textual subdivisions passed down from ancient times, or if it was instead a reflection of specific individual copying rhythms, or perhaps a consequence of the need to distribute labour among a number of scribes and illuminators, as was apparently the case with Latin Evangeliaries, according to Bischoff. In any event, the similarity of the treatment reserved for the evangelical text in the Latin and Greek contexts is striking and merits further investigation. More generally speaking, the Bible, owing its sacred nature—and status as a Book/book *par excellence*—was shared by multiple cultures, even if it was regarded differently in each of them, and therefore represents particularly fertile terrain for comparative codicological research. See Maniaci et al. 1996.

**123** In fact, in the entire corpus, a total of 435 split words at the end of quires were counted, as opposed to 4,414 at the end of pages in the *verso/recto* situation (included among which the last pages of quires) and 3,761 in the *recto/verso* situation. Since we are dealing with quaternions, there is a 1:16 chance of finding a split word at the end of a quire compared to that of finding a split word at the end of a page, and a 1:8 chance if one considers the *recto/verso*, *verso/recto* transitions separately. A comparison with two theoretical values (552 and 470 respectively)—obtained, in both cases, by dividing the frequencies surveyed in the corpus by eight—yields  $\chi^2$  values of 51.5 (very high) and 5.2, respectively. The second value, even if it is much lower than the first, is significant, with a probability of <5%.

à-vis the imbalance between the *recto/verso* and *verso/recto* instances, since it is in line with the general tendencies. In other words, the number of split words occurring at the end of quires varies in relation to the same characteristics of the codex that exerted an influence on the distribution of split words at the end of lines and pages, and it swells with the increase in split words situated in these positions. Whatever the underlying reason may have been, in this case, too, we are dealing with a widely disseminated rule, and one which was respected all the more so when the scribe, largely speaking, was particularly concerned with limiting the occurrence of split words.<sup>124</sup>

## 8 Conclusions

What overall consequences can be seen to result from the various phenomena brought to light up to this point, then? Above all, with the production of Byzantine minuscule manuscripts, starting from the earliest times, a widespread concern emerges—even if on the whole not a very pronounced one—with respect to the problems arising from the division of words located at the end of lines (i.e. ‘word splitting’). It is not unlikely that this fact constitutes evidence, albeit not of the most glaring kind, of a more general and multi-faceted intention vis-à-vis ‘line management’, which is to say a combination of strategies consciously aimed at fulfilling a dual objective, namely to maintain a number of fundamental aesthetic standards—such as the regularity of the layout and writing area—and at the same time to enhance the visual impact of the graphic ensemble and ensure recognition of its individual elements, with the ultimate goal of facilitating a more accurate and effortless scanning of the text by the reader.

Following a closer analysis, the management of word splitting has been shown to adhere not merely to one rule—applied more or less systematically and with greater or lesser degrees of care when the need arose—but rather to a series of prescriptions shaped by the potential detriment that the splitting of a word might cause to the smoothness of the textual flow. Hence, the splitting of words at

---

**124** Stated more precisely, the rule that stipulates the limitation of split words at the end of quires is closely linked to that which was aimed at avoiding them as much as possible at the end of pages. In other words, in contrast to what is seen in the relationship between split words at the end of lines and of pages—the former could be overlooked in volumes in which the latter were kept under control—the link between split words at the end of pages and quires is much closer, in the sense that scribes who were unconcerned about splits occurring at the end of pages, but who were careful to limit those situated at the end of quires, did not exist.



the end of lines was avoided wherever possible, but was not categorically proscribed. The splitting of words at the end of pages was firmly discouraged—above all when the word concerned fell between the *recto* and *verso* of a page—whilst the occurrence of a split word between one quire and the next was to be avoided as much as possible.

The attention paid to limiting the occurrence of split words at various places on pages was not equally distributed across the entire spectrum of manuscript production. In fact, it varied both in relation to a volume's material, technical and graphic characteristics, and according to its content, date and the particular circumstances of its commissioning and production. All these characteristics often prove to be associated, and not attributable—on the contrary, they are for the most part unrelated—to a fundamental opposition between high quality and more run-of-the-mill production.

Far from being the prerogative of a distinct group of scribes who had an increased awareness of the problem, as opposed to others who were entirely ignorant of it, knowledge of the rules aimed at limiting the division of words at the end of lines was widely distributed and freely available to all. On the other hand, the application of such rules was normally a matter of choice, in the sense that it was possible to respect, within certain limitations, one rule but not another, and that the use of one or all of them could largely speaking be systematically applied. Therefore, the attitude of scribes varied according to the degree of importance attributed to a problem whose solution required an undeniable expenditure of effort in relation to other issues that called for a similar amount of time and effort. This explains why the limitation of split words was more common in certain types of texts and codices—such as secular or heavily abbreviated ones—which were clearly produced by, or in relation to, a more erudite readership, or in a particularly high quality and standardised production type—like the majority of Tetra-Gospels—in which all quality and functionality requirements had to be satisfied to the maximum extent possible.

The overall picture thus described, even if it is fairly clear in terms of general trends, is perforce incomplete. The lack of a greater and more detailed knowledge of the ways in which the actors in Byzantine manuscript production operated prevents us from gaining a closer understanding of the scribes' working practices, and from being able to grasp all the factors and the rationale lying behind the choices they made. In particular, still unclear to us are the specific reasons which determined, in the management of word splitting, the contradistinctions between groups of manuscripts of different content, date and layout etc., which give rise to

statistically significant differences, albeit ones that are not equally interpretable, and about which it was decided not to furnish a more detailed report.<sup>125</sup>

Finally, it only remains to put forward the proposal that a comparison be made with the solutions developed to solve precisely the same problem in other cultural contexts—above all in the Latin setting—for which the surveys conducted up till now have already revealed the existence of a similar awareness of the problem, but the adoption of somewhat divergent attitudes and the application of different criteria for its resolution.

## References

- Beit-Arié, Malachi (1992), *Hebrew Manuscripts of East and West. Toward a Comparative Codicology*, London: The British Library (The Panizzi Lectures).
- Bergeron, Réjean / Ornato, Ezio (1990), 'La lisibilité dans les manuscrits et les imprimés à la fin du Moyen Âge', in *Scrittura e civiltà*, 14: 151-198.
- Bischoff, Frank M. (1994), 'Systematische Lagenbrüche: Kodikologische Untersuchungen zur Herstellung und zum Aufbau mittelalterlicher Evangeliare', in Rück, Peter / Boghardt, Martin (eds), *Rationalisierung der Buchherstellung im Mittelalter und in der frühen Neuzeit*. Ergebnisse eines buchgeschichtlichen Seminars der Herzog August Bibliothek Wolfenbüttel, 12<sup>th</sup>–14<sup>th</sup> November, 1990, Marburg: Institut für Historische Hilfswissenschaften (Elementa diplomatica, 2), 83–110.
- \*Bischoff, Frank M. (1996), 'Le rythme du scribe. Analyse sérielle de la densité d'écriture dans les évangiles d'Henri le Lion', in *Histoire et mesure*, 11, 1–2: 53–91.
- Blalock, Hubert M. (1984<sup>2</sup>), *Statistica per la ricerca sociale*, Bologna: Il Mulino.
- Bozzolo, Carla / Coq, Dominique / Muzerelle, Denis / Ornato, Ezio (1984), 'Noir et blanc. Premiers résultats d'une enquête sur la mise en page dans le livre médiéval', in Questa, Cesare / Raffaelli, Renato (eds), *Il libro e il testo*. Atti del convegno internazionale (Urbino, settembre 1982), Urbino 195–221 (repr. in [Ornato, Ezio et al.] [1997], *La face cachée du livre médiéval. L'histoire du livre vue par Ezio Ornato, ses amis et ses collègues. Avec une préface d'Armando Petrucci*, Roma; Viella [I libri di Viella, 10], 473–508).
- Bozzolo, Carla / Coq, Dominique / Muzerelle, Denis / Ornato, Ezio (1987), 'Page savante, page vulgaire : étude comparative de la mise en page des livres en latin et en français écrits ou imprimés en France au XV<sup>e</sup> siècle', in Baumgartner, Emmanuèle / Boulestreau, Nicole (eds), *La présentation du livre*. Actes du Colloque, Paris X-Nanterre (4, 5, 6 décembre 1985), Paris: Centre de recherches du Département de français de Paris X-Nanterre (Littérales, 2), 121–133 (repr. in [Ornato, Ezio et al.] [1997], *La face cachée du livre médiéval. L'histoire du livre vue par Ezio Ornato, ses amis et ses collègues. Avec une préface d'Armando Petrucci*, Roma: Viella [I libri di Viella, 10], 509–517).

---

**125** For example, the lack of any clue regarding the provenance of the manuscripts, which is notoriously difficult to ascertain in the absence of explicit references or sufficiently indicative external evidence, should be noted.

- Bozzolo, Carla / Ornato, Ezio (1980), *Pour une histoire du livre manuscrit au Moyen Âge. Trois essais de codicologie quantitative*, Paris: CNRS (Publications de l'équipe de recherche sur l'humanisme français des XIV<sup>e</sup> et XV<sup>e</sup> siècles. Textes et études, 2).
- Canart, Paul (1990), 'La minuscule grecque et son ductus du IX<sup>e</sup> au XVI<sup>e</sup> siècle', in Irigoien, Jean / Poulle, Emmanuel / Sirat, Colette (eds), *L'écriture: le cerveau, l'œil, la main*. Actes du colloque international du Centre National de la Recherche Scientifique, Paris, Collège de France, le 2, 3 et 4 mai 1988, Turnhout: Brepols (Bibliologia, 10), 307–320.
- Canart, Paul / Perria, Lydia (1991), 'Les écritures livresques des XI<sup>e</sup> et XII<sup>e</sup> siècles', in Harlfinger, Dieter / Prato, Giancarlo (eds), *Paleografia e codicologia greca*. Atti del II colloquio internazionale (Berlin-Wolfenbüttel, 17–21 October, 1983), with the collaboration of D'Agostino, Marco / Doda, Alberto, Alessandria: Edizioni dell'Orso (Biblioteca di Scrittura e civiltà, 3), 67–118.
- Cavallo, Guglielmo / Chartier, Roger (eds) (1995), *Storia della lettura nel mondo occidentale*, Roma-Bari: Laterza.
- Fioravanti, Giorgio (1993), *Il dizionario del grafico*, Bologna: Zanichelli.
- Gilissen, Léon (1973), *L'expertise des écritures médiévales. Recherche d'une méthode avec application à un manuscrit du XI<sup>e</sup> siècle: le lectionnaire de Lobbes, codex Bruxellensis 18018*, Gand: Éditions Story-Scientia.
- Gilissen, Léon (1977), *Prolégomènes à la codicologie, Recherches sur la construction des cahiers et la mise en page des manuscrits médiévaux*, Gand: Éditions scientifiques Story-Scientia.
- Giusti, Franco (1990), *Introduzione alla statistica*, Torino: Loescher.
- Guarducci, Margherita (1987), *L'epigrafia greca dalle origini al tardo impero*, Roma: Istituto poligrafico e Zecca dello Stato.
- Hunger, Herbert (1954), 'Die Perlschrift, eine Stilrichtung der griechischen Buchschrift des 11. Jahrhunderts', in Hunger, Herbert (ed.), *Studien zur griechischen Paläographie*, Wien: Hölzl (Biblosschriften, 5), 22–32 (repr. in Hunger, Herbert [1973], *Byzantinische Grundlagenforschung*, London: Variorum Reprints).
- Irigoien, Jean (1992), 'Manuscrit imposé ou copie sur carnets?', in *Scriptorium*, 46: 88–90.
- Lemaire, Jacques (1989), *Introduction à la codicologie*, Louvain-la-Neuve: Institut d'études médiévales (Textes, études, congrès, 9).
- Maniaci, Marilena (1995), 'Ricette di costruzione della pagina nei manoscritti greci e latini', in *Scriptorium*, 69: 16–41.
- Maniaci, Marilena (1996), 'Un repertorio da leggere fra le righe?', in *Gazette du livre médiéval*, 28: 13–22.
- Maniaci, Marilena (1996 [1998?]), *Terminologia del libro manoscritto*, Milano, Roma: Istituto centrale per la patologia del libro.
- Maniaci, Marilena (2002), *Costruzione e gestione della pagina nel manoscritto bizantino*, Cassino: Università degli studi di Cassino.
- Maniaci, Marilena / Muzerelle, Denis / Ornato, Ezio (1999), 'Une bible ... mais encore? Le portrait des manuscrits bibliques dans la catalographie moderne', in Schlusemann, Rita / Hermans, Jos / Hoogvliet, Margriet (eds), *Sources for the History of Medieval Books and Libraries*, Groningen: E. Forsten, 291–309.
- Maniaci, Marilena / Ornato, Ezio (1993), 'Che fare del proprio corpus? I. Costituzione e descrizione di una popolazione di libri a fini statistici', in *Gazette du livre médiéval*, 22: 27–37 (repr. in [Ornato, Ezio et al.] [1997], *La face cachée du livre médiéval. L'histoire du livre vue par Ezio Ornato, ses amis et ses collègues. Avec une préface d'Armando Petrucci*, Roma: Viella [I libri di Viella, 10], 67–75).

- Maniaci, Marilena / Ornato, Ezio (1995), 'Intorno al testo : il ruolo dei margini nell'impaginazione dei manoscritti greci e latini', in *Nuovi Annali della Scuola speciale per archivisti e bibliotecari*, 9: 175–194.
- Martellotti, Guido (ed.) (1964), F. Petrarca, *De viris illustribus*, I, Firenze: Sansoni.
- Mastruzzo, Antonio (1995), 'Ductus, corsività, storia della scrittura: alcune considerazioni', in *Scrittura e civiltà*, 19: 403–464.
- Muzerelle, Denis (1985), *Vocabulaire codicologique. Répertoire méthodique des termes français relatifs aux manuscrits*, Paris: CEMI (Rubricae. Histoire du livre et des textes, 1).
- Muzerelle, Denis / Ornato, Ezio (1997), 'Une affaire bien mal réglée ... contribution au débat sur le codage des schémas de réglure des manuscrits grecs', in *Gazette du livre médiéval*, 30: 26–36.
- Ornato, Ezio (1975), 'Statistique et paléographie: peut-on utiliser le rapport modulaire dans l'expertise des écritures médiévales?', in *Scriptorium*, 29: 143–156.
- Ornato, Ezio (1994), 'Exigences fonctionnelles, contraintes matérielles et pratiques traditionnelles dans le livre médiéval : quelques réflexions', in Rück, Peter / Boghardt, Martin (eds), *Rationalisierung der Buchherstellung im Mittelalter und in der frühen Neuzeit. Ergebnisse eines buchgeschichtlichen Seminars der Herzog August Bibliothek Wolfenbüttel (12–14 November 1990)*, Marburg: Institut für Historische Hilfswissenschaften, 7–31 (repr. in [Ornato, Ezio et al.] [1997], *La face cachée du livre médiéval. L'histoire du livre vue par Ezio Ornato, ses amis et ses collègues. Avec une préface d'Armando Petrucci*, Roma: [I libri di Viella, 10], 117–159).
- Parker, Malcolm B. (1987), 'The Contribution of Insular Scribes of the Seventh and Eighth Centuries to the "Grammar of Legibility"', in Maierù, Alfonso, *Grafia e interpunzione del latino nel medioevo*. Atti del Seminario internazionale (Roma, 27–29 settembre 1984), Roma: Edizioni dell'Ateneo, 1987 (Lessico intellettuale europeo, 41), 15–29 (repr. in Parkes, Malcom B. [1991], *Scribes, Scripts and Readers. Studies in the Presentation and Dissemination of Medieval Texts*, London: Hambledon Press 1991, 1–18).
- Parker, Malcolm B. (1992), *Pause and Effect: An Introduction to the History of Punctuation in the West*, Aldershot: Scolar Press.
- Parker, Malcolm B. (1995), 'Leggere, scrivere, interpretare il testo: pratiche monastiche nell'alto medioevo', in Cavallo, Guglielmo / Chartier, Roger (eds) (1995), *Storia della lettura nel mondo occidentale*, Roma-Bari: Laterza, 71–90.
- Petitmengin, Pierre (1985), 'Les plus anciens manuscrits de la Bible latine', in Fontaine, Jacques / Pietri, Charles (eds), *Le monde latin antique et la Bible*, Paris: Beauchesne (Bibles de tous les temps, 2), 89–123.
- Petrucci, Armando (1968), 'Censimento dei codici dei secoli IX–XII, Istruzioni per la datazione', in *Studi medievali* Ser. 3a, 9, 1115–1126.
- Petrucci, Armando (1984), 'Lire au Moyen Âge', in *Mélanges de l'Ecole française de Rome. Moyen Âge, temps modernes*, 96:, 603–616.
- Ruffini, S. (1996), 'Tipologia e funzioni del segno di a capo nella minuscola libraria greca dell'età media, a graduate thesis' in *Manuscript Conservation at the Scuola Speciale per Archivisti e Bibliotecari at Rome's "Sapienza" University*, supervised by Lidia Perria, and later summarised in *Nuovi Annali della Scuola Speciale per Archivisti e Bibliotecari*, 10, 1996, 304.
- Ruiz Garcia, Elisa (1988), *Manual de codicología*, Salamanca – Madrid: Fundaciôn German Sanchez Ruipérez - Pirâmide (Biblioteca del Libro. Series Maior).
- Saenger, Paul (1982), 'Silent Reading: Its Impact on Late Medieval Script and Society', in *Viator. Medieval and Renaissance Studies*, 13: 367–414.

- Saenger, Paul (1990), 'The Separation of Words and the Order of Words: the Genesis of Medieval Reading', in *Scrittura e civiltà*, 14: 49–74.
- Saenger, Paul (1990a), 'Coupure et séparation des mots sur le continent au Moyen Âge', in Martin, Henri-jean / Vézin, Jean (eds), *Mise en page et mise en texte du livre manuscrit*, Paris: Éditions du Cercle de la Librairie - Promodis, 451–455.
- Saenger, Paul (1995), *Leggere nel tardo medioevo*, in Cavallo, Guglielmo / Chartier, Roger (eds) (1995), *Storia della lettura nel mondo occidentale*, Roma-Bari: Laterza, 117–154.
- Samaran, C. (1928), an untitled paper published in the *Bulletin de la Société nationale des Antiquaires de France*, 202.
- Sautel, Jacques-Hubert (1995), *Répertoire de réglures dans les manuscrits grecs su parchemin. Base de données établie par Jacques-Hubert Sautel à l'aide du fichier Leroy et des catalogues récentes*, Turnhout: Brepols (Bibliologia, 13).
- Sautel, Jacques-Hubert (1996), 'Défense et illustration de la codification Leroy', in *Gazette du livre médiéval*, 19: 24–36.
- Supino Martini, Paola (1996), 'Scrittura e leggibilità in Italia nel secolo IX', in Scalon, Cesare (ed.), *Libri e documenti d'Italia: dai Longobardi alla rinascita delle città*. Atti del Convegno nazionale dell'Associazione Italiana Paleografi e Diplomatisti, Cividale, 5–7 ottobre 1994, Udine: Arti Grafiche Friulane (Libri e biblioteche, 4), 35–60.
- Ziegler, Konrat (ed.) (1968), *Plutarchus, Caesar*, II, 2, Leipzig: Teubner.

