Word by word, phrase by phrase, sentence by sentence:
A corpus-based study of the N₁ by N₁ construction

Per Boberg
Abstract

The present paper examines the N₁ by N₁ construction using corpus linguistic methodology. The distribution of types of the construction that occur more than once either unhyphenated or hyphenated in any subcorpus of the British National Corpus accessed through the Brigham Young University interface is examined. Written and spoken language as well as various genres are compared. Hyphenation is also investigated. A collocation analysis of some types of the construction is further carried out and it is concluded that the N₁ by N₁ construction can be part of the on a N₁ by N₁ basis construction. Results from the quantitative analysis as well as the qualitative discussion suggest that the N P N construction may be undergoing lexicalisation starting as an adverbial and moving to functioning as a premodifier. This suggestion is indicated through complementary diachronic searches in the Oxford English Dictionary. It is also indicated that the construction may follow a development pattern similar to that of N₁ to N₁. The notion of construction is discussed in relation to the N₁ by N₁ construction, and a hierarchical view of constructions is proposed as a solution to some of the problems with the term.

Keywords: construction grammar, constructions, corpus linguistics, hyphenation, institutionalisation, lexicalisation, nominalisation, N₁ P N₁, N P N, premodifier
Table of contents

1. Introduction
   1.1 Aims and research questions

2. Material and method
   2.1 The British National Corpus
   2.2 Method and scope

3. Theoretical background
   3.1. Constructions
   3.2 From collocation to idiom
   3.3 Lexicalisation

4. Results
   4.1 Written subcorpus
     4.1.1 Miscellaneous
       4.1.1.1 Collocations
       4.1.1.2 Interim summary
     4.1.2 Fiction
       4.1.2.1 Collocations
       4.1.2.2 Interim summary
     4.1.3 Academic
       4.1.3.1 Collocations
       4.1.3.2 Interim summary
     4.1.4 Newspaper
       4.1.4.1 Collocations
       4.1.4.2 Interim summary
     4.1.5 Summary and comparison of results from the written subcorpora
       4.1.5.1 Frequency and form
       4.1.5.2 Collocations
       4.1.5.3 Interim summary and conclusions
   4.2 Spoken subcorpus
     4.2.1 Collocations
   4.3 The N₁ by N₁ construction in written v. spoken subcorpora
     4.3.1 Collocations
   4.4 Complementary diachronic data

5. Discussion
   5.1 Lexicalisation
   5.2 The N₁ P N₁ construction(s)

6. Summary and conclusions

References
1. Introduction

The present paper is a mainly quantitative corpus study that investigates the \( N_1 \) P \( N_1 \) construction in cases where an identical noun precedes and follows the preposition *by*. This results in the \( N_1 \) *by* \( N_1 \) construction which can be seen in (1), compared to the \( N \) P \( N \) construction (any preposition preceded and followed by nouns) which can be seen in (2).

(1) Friday was a quick learner and his English got better **day by day**\(^1\). (FRX, \( W_{\text{fict_prose}} \))

(2) The policeman smiled back weakly. There is a **life after death**, he thought.

\( (CML, W_{\text{fict_prose}}) \)

The \( N_1 \) *by* \( N_1 \) construction has existed for a long time. A search for *by* in *The Oxford English Dictionary* (OED) Online shows the earliest occurrence in Chaucer’s *The compleynt of Venus* from circa 1392, seen in (3).

(3) To folowe **word by word** the Curiosite of Graunson. (OED)

It is described as “preceded and followed by the n. or word of quantity” (OED). Searches for other types of the construction showed even earlier occurrences.

The reason for selecting this construction as a topic of study is its specific semantic and syntactic features. The construction occurs frequently in language; two identical nouns are often conjoined by a preposition. Further, the repetition of syntactic elements within the construction makes it an interesting topic. In relation to fixedness it is interesting in the sense that the syntactic function performed by the construction may influence its fixedness; it may be subject to a process of lexicalisation. Regarding semantics, identical nouns conjoined by a preposition, i.e. instances of the construction, may for instance indicate ‘succession’ or ‘movement’ and might be prone to take on uncompositional meanings. Moreover, some instances of it seem to collocate frequently with certain nouns when it functions as a premodifier, which further indicates that construction-specific semantic mechanisms are at work. Finally, a theoretically driven motivation of the investigation of the construction is the

\(^1\) Boldface in the examples is added emphasis.

\(^2\) All examples in the present study were obtained from the Brigham Young University interface to the British National Corpus if not stated otherwise. Only section number and subcorpus is accounted for after the examples.
ambiguity of the term *construction*. Also, there seems to be a lack of clear definitions of *construction*.

Jackendoff (2008: 27) states that “the ways in which NPN deviates from canonical structures lead to the conclusion that at least some licensing of complements and modifiers is a function of semantics rather than syntax”. Lindquist & Levin (2009) examine the $N_1 \text{to} N_1$ pattern with body-part nouns and suggest that it is subject to several processes of development such as semantic layering, lexicalisation, semantic bleaching and the development of new pragmatic meanings (2009:185). Consequently, the results from these studies also suggest that processes of development are at work with the $N_1 \text{P} \ N_1$ and the $N_1 \text{to} N_1$ constructions. The present paper aims to account for the characteristics of the $N_1 \text{by} N_1$ construction(s) in order to determine if there are indications that lexicalisation may be identified for it as has been suggested for other $N_1 \text{P} \ N_1$ constructions.

1.1 Aims and research questions

The first main aim of the present paper is a descriptive one: to describe the distribution and functions of the $N_1 \text{by} N_1$ construction in spoken and written British English. This also includes collocations of some of the premodifier instances of the construction and other observed semantic features. More specifically, the following points are investigated:

- what positions the construction holds within sentences (its functions$^3$)
- what form$^4$ the construction has
- what differences there are between written and spoken language and between different written text-types$^5$ regarding the construction
- if there are tendencies for premodifier instances of the construction to collocate frequently with certain words or semantically related groups of words
- if there are signs of lexicalisation with this construction

These areas of investigation yield the first research questions:

- What are the significant characteristics of the $N_1 \text{by} N_1$ construction,

---

3 E.g. adverbials or premodifiers, the function performed within a sentence.
4 Hyphenated or unhyphenated.
5 The written part of the corpus has four main text-types (collections of genres), which constitute individual subcorpora: miscellaneous, fiction, academic and newspaper.
distributionally, syntactically and semantically by function, form\(^6\), collocation patterns, mode\(^7\) and text-type?

- Are there signs of lexicalisation with this construction?

The second main aim of the study is theoretically driven: to investigate theoretical issues that arise from the observations made about the distribution and function(s) of the construction. More specifically, the following points are discussed:

- the notion of *construction* and its applicability on the problem investigated in the present paper
- if \(N_1\) *by* \(N_1\) is a type of the \(N\ P\ N\) or \(N_1\ P\ N_1\) construction, a construction of its own, or several constructions
- what advantages and disadvantages there are of studying this/these construction(s) using corpus linguistic methodology and the construction grammar framework

These issues will be discussed from the following research question:

- Is the notion of *construction* useful in the study of \(N_1\ P\ N_1\) and is \(N_1\) *by* \(N_1\) a type of a construction, a construction of its own or several constructions?

In the following section, the material used to collect the data used in the present study is presented along with the method used to collect the data from the corpora.

2. Material and method

2.1 The British National Corpus

The corpus used for the present study is the Brigham Young University edition of the British National Corpus (henceforth, the BYU-BNC) by Mark Davies.

Leech et al. (2001) give an account of the composition and contents of the British National Corpus (henceforth, BNC). The BNC is comprised of both written and spoken data, but the written component constitutes some 90% of the total amount of corpus data.

---

6 Unhyphenated or hyphenated.
7 Spoken or written English. The fifth subcorpus contains the transcriptions of spoken language.
The written component is comprised of two “broadly defined kinds of text: imaginative […] and informative” (2001: 2). Imaginative texts are defined as fictional texts, poetry and some other texts of a literary nature. These texts constitute around 20% of the component, whereas the informative part constitutes the remaining 80%. The informative part contains non-fictional texts, which are sub-divided into eight domains: Arts (8.08%), Belief and Thought (3.40%), Commerce (7.93%), Leisure (11.13%), Natural Science (4.18%), Applied Science (8.21%) Social Science (14.80%), World Affairs (18.39%) and Unclassified (1.93%) (ibid.). These categories, in all, account for 59.66% of the total amount of data in the corpus.

The spoken component is divided into two parts: a conversational part containing spontaneous conversational interaction between individuals of 15 years or older, which contains just over 40% of the component (2001: 3), and a task-oriented part which constitutes around 60% of the component. The conversational part of the corpus is also sub-divided according to age, social group, and sex. There are also subdivisions to the task-oriented part of the spoken corpus, which contains activities that are educational and informative, of business origin, public/institutional and of leisure origin. Leech et al. (ibid.) also list subdivisions of these activity-types, such as e.g. classroom interaction for business and political speeches for public/institutional. Finally, they also present the regional distribution for the spoken component: South (45.61%), North (25.43%), Midland (23.33%) and Unclassified (5.61%) (2001:4).

The written subcorpus of the BYU-BNC is divided into four subcorpora: miscellaneous, fiction, academic and newspaper. It is important to point out that these subcorpora are not equal in size. In all, the sizes of the subcorpora of the BYU-BNC add up to 96.3 million words. The largest of these subcorpora is the miscellaneous subcorpus with 44.6 million words, followed by the fiction subcorpus with 15.9 million words and the academic subcorpus with 15.3 million words. The smallest of these subcorpora is the newspaper subcorpus with 10.5 million words. This means that the largest written subcorpus is more than four times larger than the smallest written subcorpus. Consequently, these subcorpora cannot be compared statistically; figures need to be normalised.

It is further necessary to point out that the spoken subcorpus only contains 10.38% of the total amount of corpus data, i.e., the written subcorpus is roughly nine times larger than the spoken subcorpus and that comparison between the two, consequently, cannot be done by simple quantitative comparison. Thus, figures per million words are used in the statistical

---

8 Cf. Leech et al. (2001: 3) for a complete account of these activity-types.
comparison between written and spoken as well as between the various written subcorpora. However, the most important point of that part of the investigation is to look at the qualitative differences and similarities between written and spoken, and in this statistical material can show general tendencies.

2.2 Method and scope

The present paper uses a corpus based approach to the investigation of the N₁ by N₁ construction through the study of extracted data from the BYU-BNC. Corpus data was extracted through individual searches in each BYU-BNC subcorpus using the search string [n*] by [n*] and adding the criterion of a minimum frequency of two tokens per type. For the hyphenated types, each subcorpus was searched for hyphenated instances of the construction using the search string *-by-* with the same added criterion. Consequently, all types that only occurred once, hyphenated or unhyphenated, in a subcorpus were excluded. The criterion put up for inclusion in the material is thus that a token of any given type must occur at least twice in either form in at least one of the subcorpora of the BYU-BNC. There is a possibility that a type could occur in only one instance in every subcorpus in each form, which would mean that very hypothetically, it could occur ten times throughout the corpus. These possible tokens were nevertheless disregarded in order to extract a more manageable material. The benefit of including all possible rare types did not match the exponential increase of the amount of data to analyse in that case. For instance, in the miscellaneous subcorpus, a search for unhyphenated N by N without the limitation of two tokens minimum showed that there were more than 5000 possible types of N by N.

The next step after conducting the searches was to manually remove all the tokens where by was not preceded and followed by two identical nouns, that is, all N₁ by N₂ patterns so that only the N₁ by N₁ patterns remained.

When the collected material (unhyphenated and hyphenated) was processed and sorted, that is, when the raw quantitative data had been extracted, the next step was to investigate the material with focus on the function of the construction within sentences. After this syntactic analysis, a semantic analysis followed. This was done manually, token by token, subcorpus by subcorpus. The written subcorpora were investigated in order of size, starting with the miscellaneous subcorpus, followed by the fiction, academic, and newspaper subcorpora. The written subcorpora were then compared to each other, results were summarised and compared to the spoken subcorpus. Finally, a complementary diachronic
search in the OED for the five most common types of the construction was carried out in order to find possible indications of which function(s) the construction had first.

3. Theoretical background
In this section, the theoretical background of the study is presented.

3.1. Constructions
In order to investigate the data, the \( N_1 \) by \( N_1 \) construction is regarded simply as a construction without further specification. The investigation of the data will however show whether or not the definition of \( N_1 \) by \( N_1 \) as a construction is useful and what could be considered a construction.

Jackendoff (2008) simply considered \( N \) P N to be a construction without further in-depth discussion in spite of the fact that this classification is not as straightforward as it might seem. For instance, it is not clear whether each individual node preposition in the pattern would make it a separate construction or not (i.e. what the criteria for individual constructions are), and further if other factors such as the nouns involved could give the pattern properties that would suggest regarding them as individual constructions. As mentioned, in order to have a starting point, it has been decided to regard \( N_1 \) by \( N_1 \) as a construction. However, since the definition of construction and the view of it varies, it is necessary to account for different views of the notion.

Schönefeld (2006: 2) points out that “the term (grammatical) construction has been around in studies and descriptions of language for long”. She stresses that the term has been used frequently with different meanings and presents several ways of viewing it as (starting from earlier structuralist models): “constituency on the basis of formal capacities [and] general functional features”. As a common denominator for these earlier models she states the neglect or exclusion of meaning from consideration. (2006: 12).

Newmeyer (1996: 86) gives an account of the generative linguistic view on constructions where constructions are more or less discarded as artefacts, stating that “in current generative work in the principles-and-parameters framework, the relationship between grammatical constructions and theoretical constructs is remote, so grammatical analysis fails to provide a direct description of the various structural types found in language”. Criticism against the generative linguistic view on constructions, Newmark concludes, has given rise to the framework “construction grammar”.

6
Goldberg (1995: 4) defines “a distinct construction [...] to exist if one or more of its properties are not strictly predictable from knowledge of other constructions existing in the grammar”. In other words, this means that a construction needs to have a meaning or form of its own that cannot be derived from the components contained within it, i.e. it is a “basic unit” (ibid.) of language.

In Goldberg (2006: 5) constructions are described as “learned pairings of form with semantic or discourse function, including morphemes, words, idioms, partially lexically filled and fully general phrasal patterns”. The stipulation from Goldberg (1995:4) of the unpredictability of meaning from the components as a criterion for defining something as a construction is also included. However, in her recent work, Goldberg has moved away from this unpredictability criterion, since she states that “[t]here is evidence from psycholinguistic processing that patterns are also stored if they are sufficiently frequent, even when they are fully regular instances of other constructions and thus predictable” (Goldberg, 2006:64).

Cognitive grammar is oriented towards examining the interplay between semantics and the actual expressed text/phonological form. In other words, the processes behind the “production” of language in the mind are in focus. As Langacker (1987: 56) puts it, “[c]ognitive grammar takes seriously the goal of psychological reality in linguistic description”. He defines grammatical construction as “the syntagmatic combination of morphemes and larger expressions to form progressively more elaborate symbolic structures” and states that “there is no fundamental distinction between morphological and syntactic constructions, which are fully parallel in all immediately relevant respects” (1987: 82). In Langacker’s definition, the emphasis is more on the relationship between the phonological and semantic space. The “symbolic association between a semantic and phonological structure” results in a “symbolic unit, the construct deployed in cognitive grammar for the representation of both lexical and grammatical structure” (1987: 58).

Croft and Cruse (2004: 247) argue that “[a] construction is a syntactic configuration, sometimes with one or more substantive items [...] and sometimes not”. They further describe constructions as items that have “[their] own semantic interpretation and sometimes [their] own pragmatic meaning[s]” (ibid.).

Goldberg’s (1995: 4) definition of construction, as revised by the removal of the unpredictability criterion (2006:64) is used in the present study as a preliminary definition, since cognitive processes behind the production of utterances are not in main focus. It needs
to be emphasized, however, that cognitive processes cannot be disregarded altogether since they are the basis of the production of language. Hence, the cognitive linguistic view on constructions is included in this theoretical background. Cognitive processes are inevitably connected to meaning, but as Wray (2008: 87) points out, “[s]electing one theoretical approach over another impacts on what one believes can be achieved”. In this case it is presumed that more can be achieved through preliminarily adopting Goldberg’s (1995: 4) definition of construction.

As mentioned earlier we need to pin down whether or not $N_1 by N_1$ is one construction or several constructions. Schönefeld (2006: 3) touches on this subject by discussing the “level of abstraction that a construction is associated with”. Her question is whether constructions “are to be seen as concrete” or “of a more [...] abstract character”, that is, if the actual words/expressions, or the underlying structures according to which the words are grouped, is the construction.

3.2 From collocation to idiom

The first step towards a fixed expression is that two (or more) lexical items in some way co-occur. Co-occurring lexical items are defined by Sinclair (1991: 170) as collocations:

> Collocation is the occurrence of two or more words within a short space of each other in a text. [...] Collocations can be dramatic and interesting because unexpected, or they can be important in the lexical structure of the language because of being frequently repeated. This second kind of collocation, often related to measures of statistical significance, is the one that is usually meant in linguistic discussions. [...] Collocation is a contributing factor to idiom.

(Sinclair, 1991: 170)

Sinclair (1991: 109–115) defines the open-choice principle and the idiom principle, two means of language (text) production. The open-choice principle is when words from the lexicon fill grammatical “slots” or as Sinclair puts it “a way of seeing language text as the result of a very large number of complex choices. At each point where a unit is completed [...] a large range of choice opens up and the only restraint is grammaticalness” (1991: 109). It is however pointed out that it would not be possible to “produce normal text simply by operating the open-choice principle” (1991: 110). Consequently, the idiom principle is presented as an alternative way of viewing text production: “[A] language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be reanalysable into segments” (ibid.). Sinclair (1991: 114) points out that these two models are “diametrically opposed” and “incompatible with each other”, i.e. they
never intersect – only one principle at a time is operated. It is suggested that the two principles are utilised by the producer of text and that the producer switches between them, and further, that the utilisation of the principles varies from language user to language user. It is further proposed that the idiom principle is the most frequent one when normal text is produced. “Collocation [...] illustrates the idiom principle. On some occasions words seem to be chosen in pairs or groups and these are not necessarily adjacent.” (1991: 115).

Hudson (1998: 19) describes a “cline of idiomaticity”, based on Cowie et. al. (1983), composed by four main types, from the most fixed to the freest. The first type, *pure idioms* are described as created through a process where “word combinations first establish themselves through constant re-use” and “then undergo figurative extension” to finally solidify. The next type is *figurative idioms*, idioms that have a figurative meaning but can be interpreted literally as well. These are described as less solid than the pure idioms and considered idiomatic since they are rarely subject to variation. The two types at the other end of the cline are collocations. The first collocation type is the *restricted collocation* where “one of the elements is used in a figurative sense not found outside of the collocation” whereas the other element is used in a clearly literal way. Finally, the second collocation type is the *open collocation* where the two elements are “freely recombinable and each element is used in its literal sense” (ibid.).

Moon (1998: 26) accepts the definition of collocations simply as co-occurring lexical items that are in close proximity to each other in a text, but points out that there may be terminological confusion since “*collocation* is sometimes used to designate weak kinds of [fixed expressions and idioms (FEIs)]” (1998:26). She opts to regard collocations as the “simple co-occurrence of items” but adds the term “anomalous collocation to designate a class of FEIs”. In a sense this division is only another way of describing the “cline of idiomacy” since we have the more idiomatic and “fixed” expressions at one end of the scale and collocations at the other in her model as well.

After having distinguished between *collocations* and *anomalous collocations* Moon (1998: 27) proceeds to account for various kinds of collocations: The first is “co-occurrence of co-members of the semantic fields, representing the co-occurrence of the referents in the real world”. Secondly, there are instances where words require “association with a member of a certain class or category of item[s]” and when words bear special meanings when they collocate with specific words. Thirdly, there are syntactic collocations where “a verb, adjective or nominalization requires complementation with [for instance] a specified particle”.

9
Moon (1998: 20–21) also defines different types of anomalous collocations: *ill-formed collocations* that “break the conventional grammatical rules of English”, *cranberry collocations*, which contain items that cannot be found in other collocations – unique to the string, *defective collocations* where at least “one […] of the component items is semantically depleted” or “a component item has a meaning not found in other collocations or contexts, although it has other compositional meanings” and *phraseological collocations* which Moon labels as the weakest ones. They are represented by “cases where there is a limited paradigm in operation and other analogous strings may be found, but where the structure is not fully productive” (ibid.).

In the present study, Moon’s (1998: 26) definition of collocation is used since the collocation analyses are oriented towards identifying items that co-occur with the \( N_1 \text{ by } N_1 \) construction.

### 3.3 Lexicalisation

The process of lexicalisation is a key component in the study of fixed expressions and idioms regarding how they became, or are becoming, “fixed” and how other expressions show tendencies of becoming fixed.

Moon (1998:36) describes lexicalisation as “the process by which a string of words and morphemes becomes institutionalised and develops its own specialist meaning or function”. The process of lexicalisation is relevant to the present topic since some examples of the \( N_1 \text{ by } N_1 \) pattern seem to have developed their own meaning and function. For example, as Lindquist & Levin (2009: 175) mention regarding the \( N_1 \text{ to } N_1 \) pattern, a certain string of words can carry its own meaning. “Some [constructions] could refer to movement ([…] hand to hand) others to position […] or contact”. Here, it may be the case that the structure is lexicalised and carries a meaning of its own that could not have been derived compositionally. The investigation of synchronic corpus data and a complementary search for diachronic data in the present study may indicate if the \( N_1 \text{ by } N_1 \) construction is affected by this process.

Bauer (1983: 45–61) defines the various stages towards lexicalisation in more detail. The first type of creation of a more complex word is a *nonce formation*, which is when “a new complex word [is] coined by a speaker/writer on the spur of the moment to cover some immediate need” (1983: 45). Note the difference between this term and collocation which is simply the co-occurrence of lexical items. Bauer however admits that there is disagreement as to seeing completely regular co-occurrences as *nonce formations*. The next step in Bauer’s
model is institutionalisation, whereby a nonce formation becomes “accepted by other speakers as a known lexical item” (1983: 48). In this step, Bauer states, fewer possible meanings are often used. This is exemplified by the noun telephone box, which could mean a box shaped like a telephone but does not, since a specific use is associated with the words when they occur together. Finally, Bauer describes lexicalisation, which occurs when “the lexeme has, or takes on, a form which it could not have if it had arisen by the application of productive rules” (1983: 48).

Various types of lexicalisation are also described by Bauer. For instance, semantic lexicalisation, albeit there are various classifications of it, is described as the addition or loss of semantic information (1983: 56). This classification has encountered problems in where to draw the line between it and other types of lexicalisation, but the main point that “[the] lack of semantic compositionality, i.e. the meaning of the whole is not predictable from the parts” is according to Bauer (1983: 58) a frequently discussed topic in the literature. It is however stated that lexemes can be the results of several types of lexicalisation, which according to Bauer (1983:61) is a reason why many scholars have regarded words as “lexicalized or not lexicalized, without recognizing that [they] might be lexicalized in only one way”. In the present study, Bauer’s (1983) definition of lexicalisation is used, and institutionalisation and lexicalisation are considered separate phenomena.

As mentioned earlier, Goldberg (2006: 5) presents the criterion of unpredictability of either meaning or form in order to define something as a construction. This can be related to lexicalisation in that some instances of N₁ by N₁ may have developed their own meanings. This however complicates the definition of N₁ P N₂ and N₁ P N₁ as construction(s). If one type of N₁ P N₁ develops a meaning of its own that cannot be derived compositionally, e.g. through the specific combination of words within it, it is a simplification to refer to N₁ P N₁ in general as a construction; meaning is also important.

4. Results
In the present section, results from the searches in the BYU-BNC are presented and discussed from various quantitative and qualitative perspectives⁹. It is divided into four main subsections. Firstly, results from the written subcorpus (and its subcorpora) are discussed. This is followed by a presentation of results from the spoken subcorpus and a comparison between data from the written and spoken subcorpora. Finally, the fourth subsection contains

---

⁹ For an account of these perspectives, cf. Section 1.1.
results from some complementary searches for diachronic data in the OED.

4.1 Written subcorpus

4.1.1 Miscellaneous

The present subsection presents the results from the miscellaneous written subcorpus. The miscellaneous subcorpus was the largest of the written subcorpora, with 44.6 million words, and also yielded the largest number of types of the \( N_1 \text{ by } N_1 \) construction occurring more than once, in total 68. Table 1 shows the distribution of tokens in the miscellaneous subcorpus according to type, frequency form and sentence functions.

Table 1. Distribution of the \( N_1 \text{ by } N_1 \) construction in the miscellaneous subcorpus of the BNC

<table>
<thead>
<tr>
<th>Type</th>
<th>Advl.</th>
<th>Premod.</th>
<th>Other</th>
<th>Tot.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>U</td>
<td>H</td>
<td>U</td>
<td>H</td>
</tr>
<tr>
<td>SIDE BY SIDE</td>
<td>241</td>
<td>27</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>STEP BY STEP</td>
<td>58</td>
<td>25</td>
<td>30</td>
<td>121</td>
</tr>
<tr>
<td>YEAR BY YEAR</td>
<td>56</td>
<td>1</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>DAY BY DAY</td>
<td>55</td>
<td>3</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>BIT BY BIT</td>
<td>32</td>
<td>1</td>
<td>5</td>
<td>38</td>
</tr>
<tr>
<td>CASE BY CASE</td>
<td>5</td>
<td>12</td>
<td>19</td>
<td>36</td>
</tr>
<tr>
<td>LINE BY LINE</td>
<td>19</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>MONTH BY MONTH</td>
<td>10</td>
<td>2</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>STAGE BY STAGE</td>
<td>13</td>
<td>1</td>
<td>6</td>
<td>20</td>
</tr>
<tr>
<td>WEEK BY WEEK</td>
<td>11</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>PIECE BY PIECE</td>
<td>15</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MINUTE BY MINUTE</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>ROOM BY ROOM</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>INCH BY INCH</td>
<td>11</td>
<td>2</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>COUNTRY BY COUNTRY</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>FRAME BY FRAME</td>
<td>7</td>
<td>1</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>STONE BY STONE</td>
<td>10</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SECTION BY SECTION</td>
<td>9</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>POINT BY POINT</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>HOUR BY HOUR</td>
<td>6</td>
<td>1</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>INDUSTRY BY INDUSTRY</td>
<td>4</td>
<td>3</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>ITEM BY ITEM</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>BRICK BY BRICK</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

10 Advl. = adverbial, premod. = premodifier, U = unhyphenated, H = hyphenated.
In Table 1, construction types are accounted for in three main columns according to function (adverbials, premodifiers and other). The first function, adverbial, is exemplified in (4).

(4) A built-up surround will have to be dismantled piece by piece, but with the other types, you can chop away the plaster (CCX, Instructional)

Here, we have an adverbial of manner, explaining the manner of the verb *dismantle*. The second main column shows the distribution of the premodifier function, which is exemplified in (5).

(5) Every new Singer knitting machine now sold is accompanied by the basic "step by step" video. (CK3, W_pop_lore)

As (5) shows, the noun *video* is premodified by the *step by step*-type of N₁ by N₁. Thirdly, the column labelled *other* includes items that could not be classified, or did not fit into the first two categories. One such example can be seen in (6).

(6) **WEEK BY WEEK** Week 1 The Church's Liturgy introduces us to the prophet Isaiah whose words echo throughout Advent, calling us joyfully to live in God's presence. (CCG, W_misc)
In (6), *week by week* is not part of a complete sentence, consequently its function cannot be determined or classified accurately.

Each one of these columns is subdivided into two sub-columns, showing the number of tokens of each function group that occurred unhyphenated (U) and hyphenated (H). An example of an unhyphenated instance is shown in (7).

(7) Therefore, even if two reactive fragments were *side by side* in a solid argon matrix, there would not necessarily be enough thermal energy (B7H, W_non_ac_nat_science)

The second column shows hyphenated instances of the construction, as exemplified in (8).

(8) His *country-by-country* figures --; note again his fine record against West Indies --; are as follows: (CU0, W_pop_lore)

As Table 1 shows, there is a clear dominance of the adverbial function of the construction. 746 tokens (or 68.6 per cent) of the construction functioned as adverbials compared to 333 premodifier instances (30.6 per cent) and nine other instances (0.8 per cent). 28 of the types had no premodifier tokens at all compared to ten types that had no adverbial instances.

The table further shows that unhyphenated tokens of the $N_1$ *by* $N_1$ construction were far more frequent than the hyphenated tokens. We arrive at a sum of 770 unhyphenated tokens of the $N_1$ *by* $N_1$ construction and 318 hyphenated tokens. The unhyphenated tokens are more than twice as many as the hyphenated ones.

Another general tendency can also be observed: the adverbial instances tended to be unhyphenated and the premodifier instances tended to be hyphenated. This can be seen in the summarised results at the bottom of the table, where the total of unhyphenated adverbial instances is 682 compared to 64 hyphenated adverbial instances and the total of unhyphenated premodifier instances is 82 compared to 251 hyphenated premodifier instances.

The ten most frequent types of $N_1$ *by* $N_1$ had instances of both adverbials and premodifiers. Among these ten types, adverbials were most frequent in most types, with some exceptions. *Step by step*, the second most frequent type, showed a clear dominance of premodifier instances (83 adverbials and 151 premodifiers). *Step by step* is exemplified in (9).
IN THE THIRD PART OF OUR **STEP BY STEP** SERIES, ROGER GANN EXPLAINS HOW TO PHYSICALLY CONNECT THE MACHINES TOGETHER (CTX, W_pop_lore)

*Case by case* (5 adverbials and 31 premodifiers) followed the same pattern. An example of a prenominal modifier from *case by case* is shown in (10)

(10) payment would be made to the hospital on a **case by case** basis, without any prior commitment (B2A, W_non_ac_soc_science)

*Month by month* occurred equally many times as an adverbial and a premodifier. Among the ten most common types, five, when occurring as adverbials, were only found unhyphenated. These were **day by day**, **bit by bit**, **case by case**, **stage by stage** and **week by week**.

Among the unhyphenated types, **side by side** occurred in 242 tokens, but only once as a premodifier, which suggests that something limits its occurrence in that position. The only instance where it occurred as a premodifier can be seen in (11).

(11) Circle 148 Fisons Instruments launched the new 8000 series gas chromatographs, which feature **side by side** column mounting (ALV, W_non_ac_nat_science)

**Side by side** also differed from the general tendency among the hyphenated types. There, it occurred 27 times as an adverbial and only ten times as a premodifier. An example of the adverbial function in the most predominantly adverbial-oriented type where there were instances of prenominal modifiers, **side by side**, can be seen in (12).

(12) When two guanine bases are **side by side** in the DNA chain, both can attach themselves to the same platinum atom (B74, W_non_ac_nat_science)

As we can see in (12), there is an adverbial complement after the verb. Also, the construction has a concrete meaning: denoting the position of guanine bases in a DNA chain next to each
other. Meanings of concrete position occurred frequently with *side by side*. An explanation for this might be sought in the nature of the type as a descriptor of adjacent positioning; its indication of ‘juxtaposition’ decreases the possibility for it to function as a premodifier. Regarding *side by side*, instances that indicate juxtaposition appear to be more likely to be connected to a verb than a noun. Also, they seem to be more concrete in meaning – ‘simple placement’ as exemplified in (13).

(13) All of these animals carry their pinnae **side by side** on top of the head. (FEV, W_non_ac_nat_science)

Jackendoff (2008: 10) also mentions that constructions of this type “denote juxtaposition of two entities” and “fits into a semantic paradigm with face to face”.

Finally, there is the category of unclear or other instances. As mentioned earlier, these are often titles or instances where the position of the construction could not be determined. Here, however, another function occurs in one token. This is seen in (14).

(14) If you have a lot of different dates to enter, either use a keystroke macro --; something that I will cover in a future **Step-by-Step** (but see "Further Reading") (HAC, W_pop_lore)

Here, the construction functions as a noun. It may be possible that this form has developed from the premodifier function and has been subject to a process of lexicalisation, but diachronic studies are needed to support that claim. One could suggest inserting *guide* after the construction, but instead the construction has that meaning on its own. Lindquist & Levin’s (2009: 177) results show this tendency for certain N₁ to N₁ types such as e.g. nominalised a heart-to-heart. Similar to the token in (14), their results showed that the nominalised type heart to heart was hyphenated in twelve out of fourteen instances.

4.1.1.1 Collocations

Types of the construction that displayed obvious tendencies to collocate frequently with certain words (or groups of words that have similar meanings), during a read-through of the data, were observed and are presented in the present section. Focus here is mainly on collocations of premodifier instances of the construction¹¹. The selection of types to

¹¹ If not stated otherwise, post-node collocates are discussed throughout the paper.
investigate was based on features that made them stand out from the other types, or if they displayed such features in other subcorpora. The distinguishing features are accounted for in the discussion of the collocations. The first type that was investigated was *case by case*, which can be seen in (15).

(15) Baker told Congress on May 24 that future aid would be considered on a case
by case basis\(^\text{12}\). (HL7, W_non_ac_polit_law_edu)

As we can see in (15), *case by case* collocates with *basis*. This is a frequent occurrence, which suggests that *case by case* attracts the collocate *basis* and vice versa. Out of the 31 premodifier instances of *case by case* in the subcorpus, 23 collocated with *basis*. Hence, 74.2 per cent of the instances of *case by case* collocated with *basis*. With *case by case* functioning as an adverbial, as exemplified in (16), no obvious tendencies or frequent patterns could be identified.

(16) along with memories of a series of strikes which Rocard attempted to settle case
by case in the autumn of 1988 (HKX, W_non_ac_polit_law_edu)

The strong attraction between *case by case* and *basis* when the construction is a premodifier may be explained by the fact that *basis* is used less frequently as a subject in a sentence with an adverbial N\(_1\) by N\(_1\) construction, as can be seen in (17) and (18). No such instances could be found in any of the subcorpora.

(17) ? The *basis* is *case by case*.
(18) ? The *basis* is *year by year*.

Among the unhyphenated tokens, *approach* was also a frequent collocate of *case by case* with three of the remaining four premodifier tokens. The same pattern could not be observed among the hyphenated tokens, where, apart from *basis, case by case* collocated with: *approval, wage settlements, evaluation* and *approach*. What all these collocates have in common is that they, together with the construction functioning as a modifier, describe the manner or sequence in which something is conducted. This suggests that these types of words

\(^{12}\) Emphasis added to indicate collocate. Henceforth, all underlining in examples is added emphasis to indicate collocates.
attract case by case and vice versa.

Certain collocates were more frequent than others with step by step as well, as exemplified in (19).

(19) Everybody here has studied U2’s success and come up with a step by step guide of how to make it in the music business. (ACN, W_pop_lore)

For step by step the collocate guide occurs in ten of the 30 prenominal unhyphenated tokens of this type and beginners guide in one. There are also other collocates with similar meanings, i.e. collocates that indicate ‘instruction’. Instruction, as shown in (20) occurred three times as a collocate of step by step.

(20) This competitively priced model offers three electrode input, step by step instructions on LCD graphics display with the option of five languages[.] (B0M, W_non_ac_nat_science)

Another collocate that occurred more than once was approach which was identified in two unhyphenated premodifier tokens. (21) exemplifies this token in context.

(21) The Engineering Employers Federation says that [...]" the step by step approach has been seen by all to have worked successfully and it is right that it should continue." (HHX, W_hansard)

Other collocates that in context with the construction indicated ‘instruction’ were: photos, video, examples, guide, tutorials, spreadsheets, programme, plans, routine and procedures. Consequently, when summing up, 22 out of the 30 unhyphenated premodifier tokens indicated ‘instruction’.

The hyphenated premodifier tokens of step by step were also investigated in order to see if the same tendencies were present. Due to the large number of tokens with this type when hyphenated, collocates that occur more than once are presented in Table 2.
Table 2. Collocates of hyphenated premodifier instances of *step by step*\(^{13}\)

<table>
<thead>
<tr>
<th>Collocate</th>
<th>Frequency</th>
<th>%</th>
<th>Collocate</th>
<th>Frequency</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>GUIDE</td>
<td>31</td>
<td>35.6</td>
<td>DIAGRAM</td>
<td>2</td>
<td>2.3</td>
</tr>
<tr>
<td>INSTRUCTION</td>
<td>10</td>
<td>11.5</td>
<td>DRAWING</td>
<td>2</td>
<td>2.3</td>
</tr>
<tr>
<td>APPROACH</td>
<td>8</td>
<td>9.1</td>
<td>ILLUSTRATED INSTRUCTION</td>
<td>2</td>
<td>2.3</td>
</tr>
<tr>
<td>CHANGE</td>
<td>6</td>
<td>6.9</td>
<td>ILLUSTRATION</td>
<td>2</td>
<td>2.3</td>
</tr>
<tr>
<td>PROCESS</td>
<td>5</td>
<td>5.7</td>
<td>PHOTOGRAPH</td>
<td>2</td>
<td>2.3</td>
</tr>
<tr>
<td>GUIDANCE</td>
<td>3</td>
<td>3.4</td>
<td>PROCEDURE</td>
<td>2</td>
<td>2.3</td>
</tr>
<tr>
<td>PICTURE</td>
<td>3</td>
<td>3.4</td>
<td>SUGGESTION</td>
<td>2</td>
<td>2.3</td>
</tr>
<tr>
<td>PROGRAMME</td>
<td>3</td>
<td>3.4</td>
<td>TRANSFORMATION</td>
<td>2</td>
<td>2.3</td>
</tr>
<tr>
<td>DEMONSTRATION</td>
<td>2</td>
<td>2.3</td>
<td>TOTAL</td>
<td>87</td>
<td>102</td>
</tr>
</tbody>
</table>

In total, there were 121 hyphenated premodifier instances and out of these, 87 had collocates that occurred more than once. As Table 2 shows the most common collocate is *guide*, which is more than three times more frequent than the second most common collocate. The number of frequent collocates that indicate ‘instruction’ is similarly to the unhyphenated instances also high: *guide, instruction, guidance, picture, programme, demonstration, diagram, drawing, illustrated instruction, illustration, photograph* and *suggestion* are all examples of this. It can be suggested that another grouping of collocates can be read from Table 2: apart from ‘instruction’ several collocates indicate ‘progress’; *change, process, procedure* and *transformation* are examples of this.

This type of N\(_1\) by N\(_1\) construction shares some semantic features with *case by case*. The manner or sequence in which something is conducted is also a salient feature here, which is further indicated by other collocates such as *approach, routine, instructions* and *examples*.

*Side by side* occurred eleven times as a premodifier. No obvious pattern could be revealed among the kind of words collocating with *side by side* apart from the fact that some of them were technical terminology. Collocates were: *containers, aircraft, kitplanes, position, use, twin-rotor, seating, columns* and *column mounting*. The example that collocated with *seating* had special characteristics that require discussion. Thus, it can be seen in (22).

(22) a totally new fuselage having been designed to give **side-by-side** instead of tandem *seating* and first flight is believed to have been in 1967. (CLU, W_misc)

In (22), the noun that is being premodified is not present since it is given later in the sentence.

\(^{13}\) These are accounted for as lemmas when applicable. Only collocates that occur more than once.
Quirk et. al. (1981: 422) refer to this phenomenon of non-present nouns as ellipsis; *side-by-side* is elliptical for *side-by-side seating*. Since the construction premodifies a non-present noun in ellipsis it however is still classified as a premodifier.

As discussed above, *basis* was also a frequent collocate of *case by case*. A search for \([n^*] \text{ by } [n^*] \text{ basis}\) in the BYU-BNC revealed that *basis* was also a collocate of several unhyphenated types of the construction that only occurred once. These were *stock by stock, stage by stage, slide by slide, site by site, office by office, line by line, lift by lift, farm by farm* and *contract by contract*. One of these is exemplified in (23).

(23) All key figures were live --; up to the instant the screen was called --; and available in total and on an office by office basis. (CBX, W_Commerce)

A further finding was made regarding the cases where the unhyphenated construction collocated with *basis*. In all instances, the two pre-node positions contained *on a/an*. Consequently, the \(N_1 \text{ by } N_1\) construction only collocated with *basis* in the following pattern: *on a/an X by X basis*.

Regarding the hyphenated types of the construction, we can see the same tendency. *Basis* occurs as a collocate of various premodifier instances of the construction, as mentioned earlier in its highest frequency of co-occurrence with *case-by-case*. The pre-node collocates *on a/an* always occurred when \(N_1 \text{ by } N_1\) collocated with *basis*. A search for *on * *-by-* basis* yielded three tokens fewer than a search for *-*by-* basis*. These “missing” tokens were however all subject to adjective insertion, which resulted in instances like (24).

(24) such a complex, coordinated and subtle process is beyond biological explanation on a simple step-by-step basis. (J52, W_non_ac_nat_science)

As (24) shows, the adjective *simple* merely adds some semantic information but does not constitute a necessary syntactic element, it fills an optional slot in the *on a/an X N_1 by N_1 basis* pattern. The other two tokens that did not occur using the latter search string were one instance of *case-by-case basis*, where *generous* was inserted in the slot, and the other one was *place-by-place basis*, where *individual* was inserted.

Since *guide* was a highly frequent collocate of *step by step*, the same search as for
basis was conducted with guide, using the search string [n*] by [n*] guide. This, however, yielded fewer results; apart from step by step, only two instances of temporal N₁ by N₁ constructions collocated with guide: week by week and day by day. Fewer types of the construction that collocated with guide was also shown with the hyphenated instances. Four hyphenated types of the construction that only occurred once collocated with guide.

The strong attraction between the construction and these collocates suggest that a higher degree of fixedness may exist here. It might be possible that a step by step guide and a case by case basis are retrieved from memory as single units rather than put together compositionally. Strong collocates like these may be regarded as individual constructions with possible empty slots. As the results from the searches on combinations with basis, hyphenated or not, showed, all instances of the N₁ by N₁ construction collocated with on a/an and basis (with a few instances involving adjective insertion). Given these strong collocates and their seemingly uniform fixedness to a certain pattern it is possible to discuss them as the on a/an X N₁ by N₁ basis construction.

Blow by blow is also mentioned briefly here, since it was only found hyphenated and only collocated with the noun account. An example of this can be seen in (25).

(25) He saw them at their home, talked with them, wrote to them, received letters from them --; and all the time heard blow-by-blow accounts of their sexual activities (J0W, W_biology)

Inch by inch was also investigated since it had some distinguishing features in other subcorpora. In this subcorpus it occurred only as an adverbial when unhyphenated, and only as a premodifier when hyphenated. Here, it collocated with path and steps in contexts that indicated in one case meticulous exactness, and, in the other, extremely slow progress, as exemplified in (26).

(26) The ox walk involved each individual proceeding to the podium to cast a vote by taking agonizingly slow, inch-by-inch steps. (HLL, W_non_ac_polit_law_edu)

4.1.1.2 Interim summary
In subsection 4.1.1 the N₁ by N₁ construction, its distribution and function in the miscellaneous subcorpus of the BYU-BNC is discussed. Regarding statistical findings it was
shown that the adverbial function occurred much more frequently than the premodifier function (68.6 per cent compared to 30.6 per cent) and that instances in the category other were of marginal occurrence. It was also demonstrated that hyphenation was not as frequent as unhyphenated instances and that there was a connection between sentence function and hyphenation. Adverbials tended to be unhyphenated while premodifiers tended to be hyphenated.

It was shown that the ten most frequent types occurred both as adverbials and premodifiers but that there was a significant difference in use regarding step by step which was predominantly used as a premodifier. Case by case also displayed this feature. Side by side was shown to occur as an adverbial to a high extent, even among hyphenated instances where premodifiers were more common. One instance where the construction occurred as a noun was also identified – a step-by-step.

Finally, the subsection 4.1.1.1 consisted of a collocation analysis of construction types that showed a tendency to collocate frequently with specific words. Case by case was investigated, and it was found that it collocated frequently with basis, and that in all instances where it collocated with basis it collocated with on a/an or on a/an [Adj]. A search was also conducted for other types of the construction that collocated with basis, and in all those cases they also collocated with on a/an or on a/an [Adj] Hence, it was suggested that on a/an [X] N₁ by N₁ basis may be regarded as a construction of its own. A collocation analysis was also carried out on step by step, and it showed that guide was a frequent collocate. Fewer different types of the construction collocated with guide altogether, with the hyphenated version being more productive. No pattern similar to the on a/an [X] X by X basis pattern could however be identified with guide. At the end of the subsection, blow by blow was accounted for as only occurring hyphenated and only collocating with account.

4.1.2 Fiction
In the present section and subsections, results from the Fiction subcorpus are presented. In the second largest of the subcorpora in the BYU-BNC, the 15.9 million word fiction subcorpus, there were 29 types of the N₁ by N₁ construction, as can be seen in Table 3.

As Table 3 shows, the rate of premodifiers among the instances is extremely low, only ten instances (or 2.6%) could be identified. This is a significant variation from the occurrence in the miscellaneous subcorpus, where 30.6 per cent of the instances were premodifiers. This suggests that there is a difference in the sentence function of N₁ by N₁ constructions.
depending on text type/genre. One possible explanation of this is that narrative texts from the genre fiction might tend to favour the $N_1$ by $N_1$ construction in an adverbial position. All but one of the premodifier instances occurred hyphenated.

Table 3. Distribution of the $N_1$ by $N_1$ construction in the fiction subcorpus of the BNC

<table>
<thead>
<tr>
<th>Type</th>
<th>Advl.</th>
<th>Premod.</th>
<th>Other</th>
<th>Tot.</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>H</td>
<td>U</td>
<td>H</td>
<td>U</td>
</tr>
<tr>
<td>SIDE BY SIDE</td>
<td>178</td>
<td>8</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>BIT BY BIT</td>
<td>31</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INCH BY INCH</td>
<td>28</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STEP BY STEP</td>
<td>27</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAY BY DAY</td>
<td>19</td>
<td>2</td>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td>MOMENT BY MOMENT</td>
<td>10</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PIECE BY PIECE</td>
<td>9</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>YEAR BY YEAR</td>
<td>7</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BLOW BY BLOW</td>
<td>6</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRICK BY BRICK</td>
<td>6</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOUR BY HOUR</td>
<td>6</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SECOND BY SECOND</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MINUTE BY MINUTE</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>STONE BY STONE</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COURSE BY COURSE</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>369</td>
<td>8</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>TOTAL BY FUNCTION</td>
<td>377</td>
<td>10</td>
<td>2</td>
<td>9</td>
</tr>
</tbody>
</table>

The only unhyphenated premodifier instance of the type was minute by minute, which can be seen in (27).

(27) Unfortunately, Fenella takes her duties terribly seriously and would have a minute by minute report of what Springsteen had been up to while I'd been away. (HTL, W_fict_prose)

Another observation was that instances of hyphenation are rarer in the fiction subcorpus than in the miscellaneous subcorpus. Unhyphenated instances accounted for 95.6% of the total number of instances in the fiction subcorpus.

Consequently, hyphenation among adverbials was low, which means that the pattern from the miscellaneous subcorpus persists in the fiction subcorpus as well even though the
rate of hyphenation among adverbials is significantly lower than in the *miscellaneous* subcorpus. A further observation is that among adverbial instances, hyphenation only occurred with the most common type *side by side*. Even though hyphenation was more common among the few premodifier instances, there was a noticeable high number of hyphenated instances with *blow by blow* (six out of the nine hyphenated premodifier instances). This construction type is exemplified in (28)

(28) I could give you a **blow-by-blow** account of how England lost to Portugal the other night. (A0U, W_fict_prose)

Regarding *blow by blow*, all instances of it in the subcorpus had in common that they occurred in a prenominal position and that there were no adverbial and unhyphenated instances of it, once again distinguishing *blow by blow* from many other types of the construction. Comparing the occurrence of *blow by blow* in the *fiction* subcorpus with that in the *miscellaneous* subcorpus it can be noted that it only occurred hyphenated in the *miscellaneous* subcorpus as well.

The most common type of the $N_1 \text{ by } N_1$ construction in the *fiction* subcorpus was *side by side* which occurred 188 times in the subcorpus. Out of these instances 186 were adverbials, one was a premodifier and one in the category *other*. Nine were hyphenated, which is around five per cent. An example of *side by side* can be seen in (29).

(29) Her fair hair was so long that it touched his hand whenever they sat **side by side**.

(CJX, W_fict_prose)

All hyphenated adverbials occurred with *side by side*. Hyphenated *side-by-side* is exemplified in (30).

(30) But the cows, packed closely **side-by-side** across the narrow lane, moved towards her, mooing and shaking their horned heads (B0B, W_fict_prose)

As can also be seen in Table 3, there is a large difference in frequency of the most frequent type and the second most frequent type *bit by bit*, which occurred 31 times.
There were some instances that were categorised as other/unclear, one is shown in (31)

(31) Proud banners flutter and wave, Shoulder to shoulder and side by side, Crimson poppies --; each a grave. (W_fict_poetry)

Since the construction occurs in a poetic context where the sentence does not follow traditional grammatical rules strictly it was not classified as an adverbial to *flutter* and *wave*, which would otherwise have been the case.

Another instance that fell into the other/unclear category can be seen in (32)

(32) He cut five lines from the Country *Day by Day* proof, then restored them, realizing that he had misread the layout sheet. (G12, W_fict_prose)

The reference in (32) is not clear since it appears that the construction is part of the name of a document.

4.1.2.1 Collocations
Since focus in the collocation analyses is on premodifiers, the heavy bias towards adverbials in the fiction subcorpus makes the present section thin. However, some observations were made about the premodifier and adverbial instances in the present section as well.

Given that *case by case* did not occur at all in the fiction subcorpus it was not possible to examine if the same connection between it and the collocate *basis* existed. A search using the search string for the unhyphenated tokens and *basis* did not yield any results. The search using the search string for hyphenated instances and *basis* only yielded one result, an instance of the type *game by game*, as can be seen in (33).

(33) but I can assure you it is all part of my master plan which I am painstakingly developing on a\textsuperscript{15} game-by-game basis. (FR9, W_fict_prose)

Even though only one instance could be found in the fiction subcorpus it supported the conclusions made based on the data from the miscellaneous subcorpus: when the N\textsubscript{1} by N\textsubscript{1} construction collocates with *basis*, it also tends to collocate with *on a/an*, making it even more

\textsuperscript{15} Pre-node collocate.
reasonable to regard on a/an N₁ by N₁ basis as a construction of its own. Basis attracts the N₁ by N₁ construction and vice versa, and N₁ by N₁ basis is preceded by on a/an.

Step by step was observed to be collocating frequently with guide in the miscellaneous subcorpus. In the fiction subcorpus, however, there were no instances of step by step as a premodifier. A search was carried out for instances of the construction that collocated with guide. Search strings [n*] by [n*] guide and *-by-* guide yielded no results, consequently making comparison with the results from the miscellaneous subcorpus impossible.

Side by side only occurred once as a premodifier in the fiction subcorpus. In that instance it occurred hyphenated and collocated with man, as can be seen in (34).

(34) Funny; he'd have thought Fergus would have been a side-by-side man. (G0A, W_fict_prose)

From context it appears that side by side man means a man that stands by someone’s side, literally or in a more abstract idiomatic sense. Since only one premodifier instance of side by side was found, no conclusions could be drawn about distinguishing features of it in the fiction subcorpus.

Blow by blow collocated with nouns that denoted ‘description’ (five collocations with account and one with description). This strong attraction between blow by blow and account suggests that the type might be processed as a single unit in memory. To definitely conclude this would however require detailed psycholinguistic and neurolinguistic experiments. It can however be concluded that a special attraction exists between this construction type and the collocate account.

Compared to the results from the miscellaneous subcorpus, inch by inch was very common unhyphenated (occurring in 28 tokens) but did not occur hyphenated at all. Among the unhyphenated tokens it did not occur as a premodifier at all.

4.1.2.2 Interim summary
The present section summarises the findings from the fiction subcorpus.

Regarding frequencies it was observed that premodifier instances of the construction were very rare and the distribution hence differed from that in the miscellaneous subcorpus. Also, the connection between hyphenation and premodifiers was shown, since hyphenated instances of the construction were rarer in the fiction subcorpus. It was further demonstrated
that there was a very low frequency of unhyphenated instances among the few premodifier instances.

*Side by side* stood out as a far more frequent type than other construction types, and it is noticeable that hyphenation only occurred with this type among the adverbial instances. Similarly to the *miscellaneous* subcorpus, *blow by blow* only occurred as a premodifier.

Finally, the collocation analysis did not yield as many results as that from the *miscellaneous* subcorpus since the number of premodifiers was low. One conclusion was made about *blow by blow*: it collocated frequently with *account* in the *fiction* subcorpus as well; five out of six instances collocated with *account*. These results are very similar to those from the *miscellaneous* subcorpus, where *blow by blow* only collocated with *account*.

The following section presents the results from the *academic* subcorpus.

### 4.1.3 Academic

The present section shows the results from the *academic* subcorpus. The *academic* subcorpus is the third largest subcorpus of the BYU-BNC with 15.3 million words. Thirty types of *N₁ by N₁* were found in this subcorpus. These can be seen in Table 4.

The first observation that can be made from Table 4 is that it follows the other subcorpora in the general tendency of distributional difference between adverbial and premodifier instances and unhyphenated and hyphenated instances. That is, that adverbials were more common than premodifiers instances of the construction and that the adverbial instances were mainly unhyphenated, while premodifier instances were mainly hyphenated.

There were 166 adverbial instances of the construction in the *academic* subcorpus, 155 of which were unhyphenated, and 94 premodifier instances, 64 of which were hyphenated. This results in a rate of 63.8 per cent adverbials.

Among the ten most frequent types of *N₁ by N₁*, five were more common as premodifiers. These types were *step by step, case by case, word by word, letter by letter* and *week by week*. Out of these, the premodifier function only exceeded by one instance for the latter two. The nine most common types occurred both as adverbials and premodifiers.

As mentioned, adverbials were more frequently unhyphenated in the *academic* subcorpus as well. *Side by side*, which had the highest number of instances, was very much biased towards the adverbial function, 77 out of 82 instances of it functioned as adverbials. In (35) the less common premodifier *side by side* is exemplified.
This kind of scheme is often known as a side by side scheme. (J6R, W_ac_polit_law_edu)

Table 4. Distribution of the \( N_1 \) by \( N_1 \) construction in the academic subcorpus of the BNC

<table>
<thead>
<tr>
<th>Type</th>
<th>Advl.</th>
<th>Premod.</th>
<th>Other</th>
<th>Tot.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>U</td>
<td>H</td>
<td>U</td>
<td>H</td>
</tr>
<tr>
<td>SIDE BY SIDE</td>
<td>71</td>
<td>6</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>STEP BY STEP</td>
<td>12</td>
<td>7</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>YEAR BY YEAR</td>
<td>16</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>CASE BY CASE</td>
<td>2</td>
<td>8</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>WORD BY WORD</td>
<td>4</td>
<td>2</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>LETTER BY LETTER</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>DAY BY DAY</td>
<td>6</td>
<td>2</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>LINE BY LINE</td>
<td>3</td>
<td>1</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>WEEK BY WEEK</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>BIT BY BIT</td>
<td>6</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOMENT BY MOMENT</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>CLAUSE BY CLAUSE</td>
<td>4</td>
<td>1</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>SECTION BY SECTION</td>
<td>4</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>SENTENCE BY SENTENCE</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BLOW BY BLOW</td>
<td>3</td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>155</td>
<td>11</td>
<td>30</td>
<td>64</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type</th>
<th>Advl.</th>
<th>Premod.</th>
<th>Other</th>
<th>Tot.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>U</td>
<td>H</td>
<td>U</td>
<td>H</td>
</tr>
<tr>
<td>EVENT BY EVENT</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>ITEM BY ITEM</td>
<td></td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>POINT BY POINT</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>STAGE BY STAGE</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>TERM BY TERM</td>
<td></td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>BUILDING BY BUILDING</td>
<td>2</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>CELL BY CELL</td>
<td></td>
<td>2</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>INDUSTRY BY INDUSTRY</td>
<td>2</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>INSTANT BY INSTANT</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>KEY BY KEY</td>
<td></td>
<td>2</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>MONTH BY MONTH</td>
<td></td>
<td>2</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>PAGE BY PAGE</td>
<td></td>
<td>2</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>SECOND BY SECOND</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>SEGMENT BY SEGMENT</td>
<td></td>
<td>2</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>QUARRY BY QUARRY</td>
<td></td>
<td>2</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td><strong>TOTAL BY FUNCTION</strong></td>
<td>166</td>
<td>94</td>
<td>30</td>
<td>64</td>
</tr>
</tbody>
</table>

The second most common type, step by step, occurred in 35 instances, twelve out of which were in the adverbial position. This makes step by step more common as a premodifier. Similar to the results from the miscellaneous subcorpus case by case was more common in the premodifier function as exemplified in (36).

That is not to say that further development on a case by case basis may not in the course of time add further grounds. (GVR, W_ac_polit_law_edu)

Case by case occurred as an adverbial twice, and in the other instances as a premodifier. Word by word occurred more frequently as premodifiers as well and is exemplified in (37).

---

16 Advl. = adverbial, premod. = premodifier, U = unhyphenated, H = hyphenated.
Above all, do not compose word-by-word or line-by-line. (GVJ, W_ac_humanities_arts)

4.1.3.1 Collocations
Similar to previous subcorpora, case by case tends to occur more frequently as a premodifier than as an adverbial. In the instances where it is a premodifier, as exemplified in (38) it collocates frequently with basis.

(38) Feminists can only challenge individual men on a case-by-case basis, and go on trying to make clear why particular usages are insulting. (CGF, W_ac_soc_science)

In the academic subcorpus, hyphenated case-by-case collocated with basis in four of the six instances (all premodifiers) and unhyphenated case by case collocated with basis in seven of the eight premodifier instances (in all there were ten instances of the type). Further, in all instances where case by case collocated with basis, it also collocated with on a, as is also shown in (38). Step by step collocated frequently with approach, as exemplified in (39).

(39) Barthes's commentary represents what he calls the step-by-step approach which affirms the text's plurality (H8V, W_ac_humanities_arts)

As in (39), step by step collocated with approach in three of the nineteen unhyphenated instances of the type (out of which seven functioned as premodifiers), and in two out of the sixteen hyphenated instances. However, step by step had other collocates that indicated similar semantics, meanings that indicated ‘method’ and ‘movement’. Among the hyphenated tokens, step by step also collocated with process, method, learning sequences and development. In the academic subcorpus, step by step did not collocate with guide at all.

As mentioned, side by side stood out since it was more frequent as an adverbial in hyphenated as well as unhyphenated instances. Out of the four premodifier instances among the unhyphenated tokens, two instances of side by side collocated with scheme. One of these can be seen in (40).
Where the head rent is based on rents received, so that the scheme is a \textit{side by side} scheme, the landlord should be satisfied (J6R, W\_ac\_polit\_law\_edu)

The other two unhyphenated instances of \textit{side by side} collocated with \textit{utterances} and \textit{rent sharing}. Among the hyphenated instances, only one functioned as a premodifier. It can be seen in (41).

(41) the overlap in time was not a collision of rival forces, rather the \textit{side-by-side running} of parallel tracks. (CKN, W\_ac\_humanities\_arts)

As mentioned earlier, the question is what makes \textit{side by side} function mainly as an adverbial, rather than a premodifier. One possible answer can be sought in the semantic properties of the construction type. Since \textit{side by side} indicates ‘placement’ or ‘position’, it seems that the construction has a tendency to add semantic information after a verb rather than modifying a noun. This is probably due to the fact that in a premodifier instance, the post-node collocate would only be determined as ‘adjacent’ but it would not be clear what it would be adjacent to.

\textit{Blow by blow} was only found among the hyphenated tokens in the \textit{academic} subcorpus. There, it collocated with \textit{account} in five out of the six instances, as can be seen in (42).

(42) The next two give a more \textit{blow-by-blow account} of mounting economic difficulties. (K8U, W\_ac\_polit\_law\_edu)

In the sixth instance, it collocated with \textit{description}. Consequently, it follows the tendencies from its occurrence in previous subcorpora, where it indicates ‘description’.

In the \textit{academic} subcorpus, \textit{inch by inch} did not occur at all.

4.1.3.2 Interim summary
The present section has accounted for the results from the \textit{academic} subcorpus. In the first subsection about frequency and form, Section 4.1.3.1, it was concluded that the \textit{academic} subcorpus shared the same tendency as the other subcorpora, that adverbial instances of the construction were superior in number to the premodifier ones. Also, it was shown that hyphenation was more frequent with premodifier instances and adverbials tended to be
unhyphenated.

Regarding the distributional characteristics of individual types of the construction, *side by side* was the most frequent type, occurring in high frequency as an adverbial but only in a few instances as a premodifier. Among the most frequent types, *word by word* stood out as one where hyphenated instances were more common than unhyphenated instances.

In the collocation analysis section, it was shown that *case by case* collocated frequently with *basis* and *on a*. *Side by side* was shown to collocate with *scheme* in two out of four unhyphenated premodifier instances. *Step by step* was also shown to collocate with *approach* in several cases.

### 4.1.4 Newspaper

This section accounts for the results from the fourth largest 10.5 million word subcorpus – *newspaper*.

In Table 5 we can see that the adverbial function of the construction is more common than the premodifier function. 72 instances or 67.9 per cent of the instances were adverbia
s

<table>
<thead>
<tr>
<th>Type</th>
<th>Advl.</th>
<th>Premod.</th>
<th>Other</th>
<th>Tot.</th>
<th>Type</th>
<th>Advl.</th>
<th>Premod.</th>
<th>Other</th>
<th>Tot.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIDE BY SIDE</td>
<td>25 U</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>31 U</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAY BY DAY</td>
<td>17</td>
<td>1</td>
<td>2</td>
<td>20</td>
<td>31 U</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STEP BY STEP</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
<td>31 U</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIT BY BIT</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td>31 U</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LINE BY LINE</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>7</td>
<td>31 U</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YEAR BY YEAR</td>
<td>6</td>
<td>1</td>
<td></td>
<td>7</td>
<td>31 U</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BALL BY BALL</td>
<td></td>
<td>4</td>
<td>4</td>
<td></td>
<td>31 U</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COUNTRY BY COUNTRY</td>
<td>3</td>
<td></td>
<td></td>
<td>3</td>
<td>31 U</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>68</td>
<td>4</td>
<td>25</td>
<td>5</td>
<td>106</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL BY FUNCTION</strong></td>
<td>72</td>
<td>29</td>
<td>5</td>
<td>106</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It can also be seen that hyphenation is rare among the adverbial instances (four out of 72 instances) and very common among premodifier instances (25 out of 29 instances). As the table further shows, there were two types of the construction that stood out as more frequent than the other ones and accounted for close to half of the total number of tokens in the

---

17 Advl. = adverbial, premod. = premodifier, U = unhyphenated, H = hyphenated.
The most frequent type was *side by side* with 31 tokens, and the second most frequent one was *day by day* with 20 tokens. *Side by side* is exemplified in (43).

(43) Canon Roy Porter argues that the idea that God suffers can exist *side by side* with the idea that He is unchanging (A5P, W_newsp_brdshht_nat_misc)

From these two types, there was a leap in size to the third most common type, *step by step* which had only eight tokens.

Among the ten most common types of the construction, four occurred more frequently as premodifiers than adverbials. Out of these, two were types with four tokens and three tokens respectively. The types *step by step* and *ball by ball* only occurred as hyphenated premodifiers with eight and four tokens respectively. Out of all the hyphenated types in the *newspaper* subcorpus, one occurred more frequently as an adverbial. *Side-by-side* occurred twice as a premodifier and three times as an adverbial. As has been shown in previous subcorpora, *side by side* tends to favour the adverbial function.

Regarding instances in the category *other* in the subcorpus, one of *side by side* was a heading, and could not be classified. The other instance of *side by side* is discussed in Section 4.1.4.1. The three unclear instances of *player by player* are from a book title, as can be seen in (44).

(44) Everton *Player by Player*, by Ivan Ponting (Published by Guinness Books, £14.99). (K97, W_newsp_other_report)

All instances of *player by player* from the category *other* were also found in the same part of the corpus, K97.

4.1.4.1 Collocations
In the *newspaper* subcorpus, the same types of the construction as in the previous subcorpora were investigated, starting with *case by case*. This type occurred twice hyphenated, in both instances with the post-node collocate *basis* and the pre-node collocates *on a*. This is exemplified in (45)
Each proposal to carry out genetic manipulation was assessed on a **case-by-case basis** by an Advisory Committee on Genetic Manipulation (A50, W_newsp_brdshft_nat_report)

As we see in (45), the pattern *on a N₁ by N₁ basis* occurs here as well, which supports the suggestion to regard this pattern as a subordinate construction to some instances of the *N₁ by N₁* construction.

Similar to *case-by-case*, *step-by-step* only occurred hyphenated in the *newspaper* subcorpus. Out of the eight instances, it collocated three times with *guide*, twice with *approach* and *instructions* respectively, and once with *sequences*. As has been shown about the previous subcorpora, *guide, approach* and *instruction* are frequent collocates to this type of the construction. The use of *step-by-step* with *approach* is exemplified in (46)

Taking the election results as his guide, he should decide on a **step-by-step approach** (HJ4, W_newsp_other_report)

*Side by side* occurred frequently in the subcorpus, but only once as a premodifier when hyphenated. In that form it collocated with *racing*, as shown in (47).

Meanwhile, the first **side-by-side racing** of the year will take place at Glasgow Green on Saturday (K5A, W_newsp_other_social)

This shows a strong connection to verbs. *Racing*, it can be argued, is constructed by addition of the suffix *-ing* to the noun or verb *race*. Since there is an agentive element in *racing*, the idea that *side by side*, due to its characteristics indicating ‘placement’ and ‘position’ favours functioning as an adverbial can be supported. The instance of *side by side* that occurred in the category *other* can be seen in (48).

**THREE side-by-side playgoers** in the Buxton Opera House dress circle followed the text in their Signet Shakespeares throughout this four-hour Lear. (A83, W_newsp_brdshft_nat_arts)
In (48) the noun *play* and the verb *go* are the basis of the agentive noun *playgoer*. This noun collocates with *side-by-side* and functions as a postmodifier to that node word. The reason for this classification is that *side-by side playgoers* does not refer to the act of going to the opera house, but indicates ‘placement’ while attending the event.

Connected to the discussion about *basis* as a collocate, it was also found that *year by year* collocated with *basis* in the only unhyphenated premodifier instance of the type, as can be seen in (49).

(49) they had a vested right to 20 per cent per annum plus returns on a year by year basis. (K2T, W_newsp_other_commerce)

As (49) shows, this instance of the construction is also part of the *on a/an [X] N₁ by N₁ basis* pattern, which has been discussed as a possible construction of its own. That the *N₁ by N₁* construction when collocating with *basis* once more is preceded by *on a/an* further supports the idea to regard this pattern as a construction of its own.

No instances of *blow by blow* were identified in the *newspaper* subcorpus, making comparison with other subcorpora impossible.

*Inch by inch* occurred in three instances among the unhyphenated results from the *newspaper* subcorpus, twice as a premodifier. Both premodifier instances collocated with the noun *search*. This is exemplified in (50).

(50) Today detectives undertook an inch by inch search of the area. (K1B, W_news_script)

This type also occurred once hyphenated as a premodifier, once again collocating with *search*, indicating ‘meticulousness’ or ‘slow progress’.

4.1.4.2 Interim summary
Subsections 4.1.4 and 4.1.4.1 have accounted for the results from the *newspaper* subcorpus. In the first subsection it was shown that *side by side* and *day by day* stood out with a significantly higher number of tokens than the other types. It was also concluded that it was very rare for the construction in its adverbial function to be hyphenated and that premodifiers tended to be hyphenated. Further, it was concluded that adverbials were far more frequent
than premodifiers. It was also shown that there were more instances of the unhypenated form than the hyphenated one.

Subsection 4.1.4.1 consisted of a collocation analysis. Case by case was shown to only collocate with basis and on a as a part of the on a/an N₁ by N₁ basis construction. It was also shown that step by step collocated with guide, approach, instructions and sequences, which was consistent with results from previous subcorpora. Side by side was shown to rarely occur as a premodifier and when it did it collocated with racing. It was suggested that side by side favours occurring as an adverbial due to its characteristics indicating ‘placement’ and ‘position’. Year by year occurred in the on a N₁ by N₁ basis pattern. Finally, inch by inch was shown to only collocate with the post-node collocate search.

4.1.5 Summary and comparison of results from the written subcorpora

This section summarises and compares results from the written subcorpora to enable comparison with the results from the spoken subcorpus and the general discussion of results in Section 5.

4.1.5.1 Frequency and form

The first part of the comparison and summary of the results from the written subcorpora deals with the distribution of tokens according to function and form. Table 6 shows the total statistical distribution of tokens from all written subcorpora.

<table>
<thead>
<tr>
<th>Type</th>
<th>Miscellaneous</th>
<th>Fiction</th>
<th>Academic</th>
<th>Newspaper</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corpus size, MW</td>
<td>44.6</td>
<td>15.9</td>
<td>15.3</td>
<td>10.5</td>
<td>86.3</td>
</tr>
<tr>
<td></td>
<td>U</td>
<td>H</td>
<td>T</td>
<td>U</td>
<td>H</td>
</tr>
<tr>
<td>Adverbials</td>
<td>682</td>
<td>64</td>
<td>746</td>
<td>369</td>
<td>8</td>
</tr>
<tr>
<td>Premodifiers</td>
<td>82</td>
<td>251</td>
<td>333</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Other</td>
<td>6</td>
<td>3</td>
<td>9</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>TOTAL</td>
<td>770</td>
<td>318</td>
<td>1088</td>
<td>372</td>
<td>17</td>
</tr>
<tr>
<td>TOTAL BY SUBCORPUS</td>
<td>1088</td>
<td>389</td>
<td>260</td>
<td>106</td>
<td></td>
</tr>
</tbody>
</table>

In order to make a proportional comparison, normalising the figures for occurrence per million words (henceforth MW) is a practical method. Table 7 shows the distribution per million words of tokens from all written subcorpora.

---

18 Million words.
Table 7. Distribution per MW of the \( N_1 \) by \( N_1 \) construction in the all written subcorpora

<table>
<thead>
<tr>
<th>Type</th>
<th>Miscellaneous</th>
<th>Fiction</th>
<th>Academic</th>
<th>Newspaper</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corpus size, MW</td>
<td>44.6</td>
<td>15.9</td>
<td>15.3</td>
<td>10.5</td>
<td>86.3</td>
</tr>
<tr>
<td>Adverbials per MW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>15.29</td>
<td>1.43</td>
<td>23.2</td>
<td>0.5</td>
<td>6.47</td>
</tr>
<tr>
<td>H</td>
<td>16.73</td>
<td>5.63</td>
<td>0.06</td>
<td>0.57</td>
<td>0.38</td>
</tr>
<tr>
<td>T</td>
<td>24.39</td>
<td>7.47</td>
<td>0.63</td>
<td>4.18</td>
<td>2.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.14</td>
<td>6.14</td>
<td>0.06</td>
<td>2.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.36</td>
<td>4.04</td>
<td>5.4</td>
</tr>
<tr>
<td>Other per MW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>0.14</td>
<td>0.07</td>
<td>0.13</td>
<td>0</td>
<td>0.48</td>
</tr>
<tr>
<td>H</td>
<td>0.2</td>
<td>0.13</td>
<td>0.13</td>
<td>0</td>
<td>0.48</td>
</tr>
<tr>
<td>T</td>
<td>24.39</td>
<td>23.4</td>
<td>10.10</td>
<td>10.1</td>
<td>21.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16.99</td>
<td>7.33</td>
<td>5.09</td>
<td>21.36</td>
</tr>
<tr>
<td>Form per MW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>17.26</td>
<td>7.13</td>
<td>23.4</td>
<td>1.07</td>
<td>10.1</td>
</tr>
<tr>
<td>H</td>
<td>24.39</td>
<td>1.07</td>
<td>12.09</td>
<td>4.9</td>
<td>16.27</td>
</tr>
<tr>
<td>T</td>
<td></td>
<td>24.47</td>
<td>16.99</td>
<td>2.76</td>
<td>5.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10.1</td>
<td></td>
</tr>
<tr>
<td>TOTAL PER MW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21.36</td>
</tr>
</tbody>
</table>

As Table 7 demonstrates, when observing the total number of tokens per million words, the corpora had more instances of the construction per MW the larger they were, with the fiction subcorpus as an exception.

The effect with more tokens per million words would be expected, since types with only one token were excluded – the larger the subcorpus, the more types with two or more tokens will occur. However, the difference in distribution is not linear, which makes it possible to draw other conclusions than the one hypothesised above. This further shows that the fiction subcorpus differs from the other subcorpora in this respect, with a large number of instances for its size.

Miscellaneous and fiction are the subcorpora with the largest number of instances of the construction per million words, 24.39 and 24.47 million words respectively. As explained earlier, the size of the miscellaneous subcorpus is a plausible explanation for the most frequent occurrence of the construction there, but there are other factors that need to be taken into consideration as well. Text from the fiction subcorpus is of a more imaginative character. This also partly applies to texts from the miscellaneous subcorpus. Further, texts classified as fiction can generally be seen as more expressive, which leads us to the conclusion that the \( N_1 \) by \( N_1 \) construction is used more in expressive/imaginative text than in specialised/non-fiction text. The results from the academic and newspaper subcorpora support this claim. There is a significant difference in frequency between these subcorpora and the miscellaneous and fiction subcorpora. The academic subcorpus contains instances of the construction at a rate of 16.99 per million words, whereas the newspaper subcorpus only has 10.10 instances per million words. The reason for the very low result in the newspaper subcorpus may be connected to the fact mentioned earlier, that there are more types of the construction with two or more tokens in the larger subcorpora, but nevertheless, the difference in occurrence is
noteworthy. If the N₁ by N₁ construction is connected to expressiveness, then the formal restraints on academic and newspaper language (less expressive and more formal text) would naturally make the N₁ by N₁ construction less frequent in these subcorpora.

Tables 6 and 7 also show that in the written subcorpus, adverbials occur 1361 times, or 15.77 times per MW, whereas premodifiers only occur 466 times or 5.4 times per MW. Adverbal instances of the construction were predominantly unhyphenated (1274 instances out of 1361 in total) whereas premodifier instances were predominantly hyphenated (349 out of 466 in total). This distributional pattern was similar in all written subcorpora. These data however showed that premodifier instances were more often unhyphenated than adverbial instances were hyphenated. This was also a pattern that persisted in all written subcorpora. Unhyphenated adverbials should therefore be considered the elementary form of the N₁ by N₁ construction.

There is some variation between the individual subcorpora. For instance, in the fiction subcorpus, adverbials accounted for 96.9 per cent of the tokens, compared to the other subcorpora where adverbials account for 63.8 up to 68.6 per cent of the tokens. The subcorpus with the lowest rate of adverbials (63.8%) compared to premodifiers was the academic subcorpus, which suggests that prenominal modification using the N₁ by N₁ construction is more common in academic language. Figure 1 illustrates this difference in distribution more clearly.

![Figure 1. Proportion of adverbial, premodifier and other instances of the construction in the written subcorpora](image)

As Figure 1 shows, the subcorpus that differs from the other subcorpora in this respect is fiction. As can be expected, since we have established a connection between the premodifier function and hyphenation, it is no surprise that the fiction subcorpus had the smallest number
of hyphenated instances, as is demonstrated more clearly in Figure 2.

Comparing Figures 1 and 2 and observing tables 6 and 7, the correlation between adverbial function and the unhyphenated instances and premodifier function and hyphenation is obvious. The instances in the category other are few and hence do not affect the general tendencies.

![Figure 2. Proportion of hyphenation in the written subcorpora](image)

Instances that were classified as parts of the category other should also be mentioned briefly. Most of these served as headlines or were part of incomplete sentences or poetry that obstructed categorisation. Since they were few they could not be shown as parts of any significant quantitative tendencies. The instance in the miscellaneous subcorpus where \( N_1 \) by \( N_1 \) functioned as a noun, shown in (14), is however of interest as it can be seen as part of a lexicalisation process. Results from the spoken subcorpus will further exemplify this (Cf. Section 4.2).

Since some types of the construction are used a lot more frequently than others, these need to be observed separately. Therefore, the five most frequent types of the construction were summarised in Table 8.

As Table 8 shows, the five most frequent types can be found in all written subcorpora in six instances or more per subcorpus. Side by side was shown to be the most common type in all subcorpora, followed by step by step, which was the second most common type in all subcorpora but the newspaper subcorpus. Step by step is interesting to observe, since it is used more as a premodifier (as collocation analyses have also confirmed) in all subcorpora except fiction.

In the newspaper subcorpus it only occurred as a premodifier. Otherwise, the adverbial function dominated the five most frequent types of the construction. 948 adverbial instances
were present compared to 241 premodifiers and 14 other instances.

Table 8. The five most frequent types of the construction in all written subcorpora

<table>
<thead>
<tr>
<th>Type</th>
<th>Miscellaneous</th>
<th>Fiction</th>
<th>Academic</th>
<th>Newspaper</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIDE BY SIDE</td>
<td>268</td>
<td>11</td>
<td>186</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>STEP BY STEP</td>
<td>83</td>
<td>151</td>
<td>7</td>
<td>27</td>
<td>12</td>
</tr>
<tr>
<td>DAY BY DAY</td>
<td>55</td>
<td>13</td>
<td>1</td>
<td>19</td>
<td>3</td>
</tr>
<tr>
<td>YEAR BY YEAR</td>
<td>57</td>
<td>13</td>
<td>1</td>
<td>16</td>
<td>3</td>
</tr>
<tr>
<td>BIT BY BIT</td>
<td>32</td>
<td>6</td>
<td>31</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

A further point can be made about the frequency of the five most common types of the construction. Alone, these account for 1203 out of the 1843 instances in total in all written subcorpora. This is roughly 65.3% of the total number of tokens. Moreover, these five most common types are also the most common types of the N by N construction (N₁ by N₁ and N₁ by N₂). Jackendoff (2008: 8) claims that “most instances of NPN require the two nouns to be identical”. Since these five most common types of NPN are highly frequent, it might be the case that Jackendoff’s claim is accurate. However, given the large number of types and the exclusion of types that only occur once it is not possible to support that conclusion for N by N with the present corpus data.

4.1.5.2 Collocations

In all four written subcorpora, case by case was investigated for its collocational patterns. In the miscellaneous subcorpus, 22 out of 30 premodifier instances of case by case collocated with basis. No conclusions regarding case by case could be drawn from the fiction subcorpus since it did not occur there at all. In the academic subcorpus, it collocated with basis in eleven out of fourteen instances. In all instances it also collocated with on a, resulting in the pattern on a case by case basis. Finally, in the newspaper subcorpus, it only occurred hyphenated, and in all instances collocated with on a and basis. In the miscellaneous subcorpus, it was also shown that case by case collocated with approach in three of the eight remaining premodifier instances. The collocation with approach was also found to be frequent in the case of the type step by step.

In the miscellaneous subcorpus, step by step collocated with guide in ten out of thirty premodifier instances of the construction and three times with the lemma instruction. It also collocated with approach twice among unhyphenated premodifier instances. It was also
demonstrated that many other less frequent collocates of the construction type indicated ‘instruction’. In all, 22 out of 30 unhyphenated premodifier instances indicated this meaning. Regarding hyphenated instances, guide was the most common collocate, with 31 instances followed by instruction with ten instances. Other collocates that indicated ‘instruction’ were also present. In the fiction subcorpus, step by step did not occur as a premodifier. Regarding the academic subcorpus, step by step collocated with approach in a few of the prenominal instances of the subcorpus. However, it did not collocate with guide at all here, which indicates genre differences. In the newspaper subcorpus, step by step only occurred hyphenated, but collocated with guide, approach and instructions. The tendency for this type to collocate with words denoting ‘instruction’ and the word guide was thus shown to exist in this subcorpus as well.

In the miscellaneous subcorpus, separate searches were made for a seemingly common collocate of the construction. The noun basis only occurred with instances of the N_i by N_i construction and these instances were shown to collocate in all cases with the pre-node collocates on a/an. This is grounds for regarding the the string N_i by N_i basis as a part of the on a/an [X] N_i P N_i basis construction. A similar search using the collocate guide was performed. This showed that step by step collocated frequently with guide. It was suggested that these instances of the construction, with collocates, could be retrieved as single units from memory rather than compositionally. In the fiction subcorpus, a search for basis only revealed one instance, which however supported the claims from the miscellaneous subcorpus.

Side by side occurred six times as a premodifier in the miscellaneous subcorpus. In the fiction subcorpus, the type only occurred once as a premodifier. In the academic subcorpus, side by side collocated with scheme in two out of four premodifier instances, and it also had the collocate racing in the premodifier function.

In the miscellaneous subcorpus, blow by blow was shown to collocate frequently with the noun account. This tendency was also shown in the fiction subcorpus (five out of six instances) and in the academic subcorpus (five out of six instances there as well). In the newspaper subcorpus no instances of blow by blow were present.

Inch by inch collocated with path and steps in the miscellaneous subcorpus when hyphenated. In the fiction subcorpus it only occurred as an adverbial. In the academic subcorpus it did not occur at all. Finally, in the newspaper subcorpus it collocated twice with
search indicating meticulousness, similar to the collocations from the miscellaneous subcorpus.

4.1.5.3 Interim summary and conclusions
Summarising the results from the written subcorpora, several conclusions can be drawn from each of the aspects investigated.

Regarding the distribution of instances in the corpora it was demonstrated that a subcorpus had more instances of the construction the larger it was, with the fiction subcorpus as an exception. It was suggested that frequency of the construction depended on genre, since more imaginative texts such as fiction had more instances of the construction and academic and newspaper had fewer.

Regarding sentence functions it was shown that the construction is used predominantly as an adverbial. The second most common function is the premodifier function. It was also shown that the construction functioning in other ways or as a noun was of marginal occurrence. The connection between the premodifier function and hyphenation was also shown for all subcorpora in that adverbials occurred more frequently unhyphenated and premodifiers occurred mostly hyphenated. A possible explanation to the connection between the premodifier function and hyphenation may be that premodifier instances are more fixed and therefore occur hyphenated more frequently than adverbials, since hyphenation seems to indicate a stronger connection between words than a simple space.

Finally, the collocation analysis revealed some differences between subcorpora in how instances of the construction collocate. Case by case collocated very frequently with basis, and it was shown that in all these cases it is a part of the on a/an [X] N₁ by N₁ basis construction. Separate searches revealed that other instances of the N₁ by N₁ construction collocating with basis also collocated with on a/an. It was further demonstrated that case by case collocated frequently with approach, similar to step by step. That construction type also collocated frequently with guide and nouns denoting ‘instruction’, the most common collocate being instruction. These tendencies were shown for all subcorpora but fiction, where the type did not occur as a premodifier.

4.2 Spoken subcorpus
The present section presents the results from the searches in the spoken subcorpus of the BYU-BNC. This subcorpus contains 10 million words and the distribution of N₁ by N₁ is
shown in Table 9.

As can be seen in Table 9, seventeen types of the N1 by N1 construction were identified in the spoken subcorpus of the BYU-BNC. The most frequent type was step by step, which occurred 25 times in the subcorpus and can be seen in (51).

(51) Growing together is a step by step process. (F86, S_meeting)

Other types with a frequency of over ten instances were day by day and side by side (17 instances respectively), year by year (15 instances) and bit by bit (14 instances).

As Table 9 shows, the step by step type of the construction occurred 25 times in this subcorpus. Predominantly, there were more adverbial instances of the construction than premodifier instances and unlike the written subcorpora, there were no types that occurred more frequently as premodifiers than adverbials.

Table 9. Distribution of the N1 by N1 construction in the spoken subcorpus of the BNC

<table>
<thead>
<tr>
<th>Type</th>
<th>Advl.</th>
<th>Premod.</th>
<th>Other</th>
<th>Tot.</th>
</tr>
</thead>
<tbody>
<tr>
<td>STEP BY STEP</td>
<td>13</td>
<td>11</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>DAY BY DAY</td>
<td>14</td>
<td>2</td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>SIDE BY SIDE</td>
<td>16</td>
<td>16</td>
<td></td>
<td>32</td>
</tr>
<tr>
<td>YEAR BY YEAR</td>
<td>13</td>
<td>2</td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td>BIT BY BIT</td>
<td>14</td>
<td>14</td>
<td></td>
<td>28</td>
</tr>
<tr>
<td>WEEK BY WEEK</td>
<td>8</td>
<td>1</td>
<td>9</td>
<td>18</td>
</tr>
<tr>
<td>LINE BY LINE</td>
<td>7</td>
<td>1</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>MONTH BY MONTH</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>PIECE BY PIECE</td>
<td>4</td>
<td>4</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>TOTAL</td>
<td>104</td>
<td>23</td>
<td>4</td>
<td>131</td>
</tr>
<tr>
<td>TOTAL BY FUNCTION</td>
<td>104</td>
<td>23</td>
<td>4</td>
<td>131</td>
</tr>
</tbody>
</table>

Among these results, the type with the highest number of premodifier instances was step by step, as exemplified in (52).

(52) Alternatively, the step by step approach urged by Mr, appears to be totally illogical. (HVH, S_pub_debate)

---

19 Advl. = adverbial, premod. = premodifier, U = unhyphenated, H = hyphenated.
Eleven out of the 25 instances of *step by step* functioned as premodifiers. This type also accounted for nearly half of the total number of premodifier instances in the subcorpus (11 out of 23). Out of all the types, only three had more than one premodifier instance, these were *month by month* (three instances), *day by day* and *year by year* (two instances respectively). With no types were premodifiers more frequent than adverbials. The only type with more than two tokens where premodifiers were as frequent as adverbials was *month by month* (three adverbial and three premodifier instances).

The category of other or unclear instances was somewhat larger than in other subcorpora per million words. This is not a surprise since spoken language naturally consists of more “ungrammatical” utterances than written language, i.e. production of speech on the fly will cause utterances that are less consistent with traditional grammatical “rules”. The occurrence of these items was however not significantly larger, which might suggest that the construction is retrieved using the idiom principle. The other instance of the *step by step* type is shown in (53)

(53) I mean, I don’t, we don’t know what the process is, of get you know, the **step by step** until it gets to the Cha Lord Chancellor, do we. (KGR, S_unclassified)

As (53) shows, *step by step* functions as a noun. From context it is clear that the noun *process* could be inserted after *step by step*, but instead *step by step* is preceded by the article and not succeeded by any noun, making it perform the nominal function. One could also see this as an interruption in the fluency of speech. However, given the statement earlier in the utterance involving the noun *process* it has been decided to classify it as a noun. The utterer may have excluded *process* to avoid repetition, thereby using *step by step* as a noun. As mentioned, *step by step* also functioned as a noun in the written miscellaneous subcorpus.

In (54) we find another example of a sentence where the construction was classified as unclear/other.

(54) Well, what we wan na do is do one, do week by week and then split it up into **day by day**. (KPN, S_conv)

In (54) *day by day* functions as a part of a prepositional phrase, namely a preposition.
complement.

There was one type that only occurred in the category Unclear/other. *Floor by floor*, which is shown in (55) and (56) only functioned as a noun.

(55) yeah and create accidents but erm well perhaps we could do it in a sort of **floor by floor** actually this Yes. (FLS, S_meeting)

(56) As, as, as somebody pointed out, I mean that the **floor by floor** was great for one, from one point of view, wasn't it? (JYL, S_unclassified)

In (55) *floor by floor* functions as a noun, preceded by the indefinite article and the non-obligatory pre-determiner *sort of*. In (56), *floor by floor* is preceded by the definite article, and not succeeded by a noun making it function as the noun of the phrase. Given the context of these instances the meaning is ‘to go through/search’ etc. one floor at a time – a **floor by floor**.

In the spoken subcorpus, three types where tokens occurred hyphenated were identified. These are not included in the data since each type only occurred once. Further, the occurrence of hyphenated instances in a spoken subcorpus does not tell us anything about hyphenation other than the preferences of the transcriber, which reduces the relevance of these instances.

4.2.1 Collocations

There were no instances of *case by case* in the spoken subcorpus. Consequently, no tendencies regarding its collocation with *basis* shown in other subcorpora could be identified.

A separate search for *[n*] *by* *[n*] *basis* in the BNC however revealed several types of the construction that occurred only once that collocated with basis. Also, among types that occurred among the original results, *year by year* and *month by month* collocated twice with *basis* respectively. For *year by year* this means that all premodifier instances of it in the spoken subcorpus collocated with *basis*, and regarding *month by month*, two out of three premodifier instances collocated with *basis*. A further finding was also made: In all cases where *N₁ by N₁* collocated with the post-node collocate *basis*, the construction also collocated with *on a*, with one exception where *erm* was inserted between *on a* and the construction, which is shown in (57).
(57) John Dennis has done a very useful spreadsheet of erm video profitability on a 
    erm video by video basis (KRY, S_meeting)

As was shown in the written subcorpora, this occurrence was frequent there as well, which 
supports the argument to regard the N₁ by N₁ construction as a part of the larger on a/an [X] 
N₁ by N₁ basis construction.

There is also the step by step type which has been shown in previous sections to 
frequently collocate with basis and approach. This tendency was also observed in the spoken 
subcorpus. Step by step is exemplified in (58)

(58) Erm we’ve heard about the measured approach, the step by step approach er and 
    Mr, has said, please make your mind up (J9S, S_pub_debate)

In (58) there is premodification of the noun approach. Eight of the eleven premodifier 
instances of the construction type in the spoken subcorpus preceded approach, this including 
one instance where it premodified measured approach. The other three instances of the 
construction type that were premodifiers collocated with process, guide and logical statement.

Side by side had no instances of prenominal modification. Two instances of blow by 
blow were identified in the present subcorpus, one of which was a premodifier. This instance 
is shown in (59).

(59) the portrait itself would be a blow by blow account of its author's story so far 
    (KRH, S_brdcast_discussn)

Similar to results from other subcorpora, blow by blow collocated with account here too. It is 
not possible to draw any general conclusions about it more than that the results correspond 
with those from other subcorpora. Finally, inch by inch did not occur in the results from the 
spoken subcorpus.

4.3 The N₁ by N₁ construction in written v. spoken subcorpora
This section compares the results from the written and spoken subcorpora in order to 
determine differences between them. In terms of form not many conclusions can be drawn 
about the difference between written and spoken English except for the lack of hyphenated
instances in spoken English. The only reason for the occurrence of hyphenation is the personal judgement and taste of the transcriber. The frequency of written and spoken instances of the construction is shown in Table 10.

Table 10. *Distribution of the N₁ by N₁ construction in spoken and written subcorpora of the BNC*

<table>
<thead>
<tr>
<th>Type</th>
<th>Written</th>
<th>Spoken</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corpus size, MW</td>
<td>86.3</td>
<td>10.0</td>
<td>96.3</td>
</tr>
<tr>
<td>Adverbials</td>
<td>1361</td>
<td>104</td>
<td>1465</td>
</tr>
<tr>
<td>Premodifiers</td>
<td>466</td>
<td>23</td>
<td>489</td>
</tr>
<tr>
<td>Other</td>
<td>16</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>1843</td>
<td>131</td>
<td>1974</td>
</tr>
</tbody>
</table>

In order to make a proportional comparison between spoken and written subcorpora the frequency per million words of written and spoken instances of the construction is shown in Table 11.

Table 11. *Distribution per MW of the N₁ by N₁ construction spoken and written subcorpora*

<table>
<thead>
<tr>
<th>Type</th>
<th>Written</th>
<th>Spoken</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corpus size, MW</td>
<td>86.3</td>
<td>10.0</td>
<td>96.3</td>
</tr>
<tr>
<td>Adverbials per MW</td>
<td>15.77</td>
<td>10.4</td>
<td>15.21</td>
</tr>
<tr>
<td>Premodifiers per MW</td>
<td>5.4</td>
<td>2.3</td>
<td>5.08</td>
</tr>
<tr>
<td>Other per MW</td>
<td>0.19</td>
<td>0.4</td>
<td>0.21</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>21.36</td>
<td>13.1</td>
<td>20.5</td>
</tr>
</tbody>
</table>

Observing tables 10 and 11, we can see that there is a clear difference in frequency between the construction in written and spoken language. N₁ by N₁ is much more frequent in written language than in spoken, occurring 21.36 times per MW in written and only 13.1 times per MW in spoken. Regarding function the use in spoken language is similar to that in written. The adverbial function is far more common than the premodifier function. In the written subcorpus, adverbials accounted for 73.8 per cent of the instances and in the spoken one they accounted for 79.4 per cent. Premodifiers followed. The category of other instances was very small in the spoken subcorpus as well. Most instances were possible to classify, and there

---

20 It should be noted that the spoken subcorpus was not compared individually with each of the written subcorpora. Since there is variation between the written subcorpora it can be concluded that the construction is more common in spoken than in *newspaper* and closer to *academic* than other written subcorpora were. The general tendency is however that the construction is more common in the written mode.
were very few instances of the construction that functioned as nouns. Three instances in the spoken subcorpus were classified as nouns, compared to only one in the written subcorpora. Step by step occurred in one spoken and one written instance as a noun. This is however too small a result to draw many general conclusions from. The other type that occurred as a noun twice was floor by floor, which only occurred in the spoken subcorpus.

The five most frequent types of the construction in the written subcorpus were side by side, step by step, day by day, year by year and bit by bit, as shown in Table 8. In the spoken subcorpus, the most frequent types were step by step, day by day, side by side, year by year and bit by bit, as shown in Table 9. It seems that the results from written are fairly consistent with those from spoken, all of the five most frequent types are the same in the written and spoken subcorpora. The difference is regarding their frequency. The most common type in the spoken subcorpus was step by step while the most common in the written one was side by side. Also, day by day was the second most common type in the spoken subcorpus compared to the third in the written one. The conclusion that can be drawn from this is that there is some difference in use of the most frequent types of the construction between the written and spoken modes, but that the five most frequent types are the same in both written and spoken subcorpora.

4.3.1 Collocations

Regarding collocations, results from the written and spoken subcorpora were compared for the identification of common features. The type case by case could not be compared since it was not identified in the spoken subcorpus. However, a search for basis, a common collocate of case by case in the written subcorpus, was carried out in the spoken subcorpus as well. It was shown that similar to the case in the written subcorpus, it was a collocate to several types of $N_1$ by $N_1$. In all cases this also collocated with on a, forming the pattern on a/an [X] $N_1$ by $N_1$ basis.

Step by step collocated with guide in many cases in the written subcorpus, which was also true for the spoken subcorpus. Another frequent collocate to step by step in both spoken and written subcorpora was approach. In all, step by step was demonstrated to indicate ‘instruction’.

Blow by blow was shown to collocate with account in the written as well as spoken subcorpus.

Summing up regarding collocations, it can be concluded that no obvious differences
between spoken and written language could be observed. The collocations that were investigated appear to be used in the same types of contexts in the written and spoken subcorpora and indicate the same meaning in applicable cases. The \( N_1 \) by \( N_1 \) construction seems to attract basis and approach and these words seem to attract the construction. The construction and basis then constitute the on a/an [X] \( N_1 \) by \( N_1 \) basis construction.

### 4.4 Complementary diachronic data

In order to substantiate whether or not there are indications that the construction may be undergoing lexicalisation, some brief searches in the OED were conducted. The five most common types of the construction were used as search strings in order to identify indications of development. Firstly, a search for side by side was conducted. The earliest instance of this type was dated to 1205 and can be seen in (60).

(60)  \( c1205 \) LAY. 19824 [a duze\(\overline{e}\)e..hine \(\overline{v}\)er bureden his bro\(\overline{d}\)er; side bi side beiene heo \(\overline{v}\)er ligge\(\overline{d}\). (OED)

The first premodifier instance of this type that could be found was however far more recent. This instance from 1908 can be seen in (61).

(61)  \( 1908 \) Daily Chron. 29 Nov. 9/5 Together with the far greater sociability afforded by the side-by-side accommodation. (OED)

For step by step, the first instance of the construction found was dated to 1580, but no premodifier instances could be found earlier than 1803. The first appearance of day by day could be found in a document from 1362, but the first premodifier instance of that type appeared in a text from 1836. The same pattern followed for year by year, which appeared in a text from 1380 but not as a premodifier in texts before 1974. Finally, bit by bit was found earliest in a text from 1716 and appeared as a premodifier in a document from 1840.

It needs to be pointed out that more extensive searches are needed in order to make diachronic conclusions, but these data indicate that the adverbial function has been present for longer than the premodifier function.

---

21 In the results from these searches spelling variations were disregarded.
5. Discussion
In the present section, theoretical issues that arose during the investigation are discussed, related to the aims and research questions stated in section 1.1.

5.1 Lexicalisation
Lindquist & Levin (2009: 176–177) describe the “paths of change” for the N₁ to N₁ pattern. Regarding that type of the N₁ P N₁ construction, “one step towards institutionalization is taken when the first preposition in the adverbial is omitted [...] turning the two conjoined prepositional phrases into a noun + prepositional phrase”. This gives us a construction such as, in their example, from hand to hand being reduced to hand to hand. Data from the present study has not shown conjoined prepositional phrases creating the construction, which might be explained by by not indicating initiated and ended movement in a sequence with another preposition such as from – to do. Rather it is more common that the construction is preceded by a noun or a verb in its adverbial function, as can be seen in (62).

(62) I live day by day, you know what I mean? (CH3, W_newsp_tabloid)

Patterns with a preceding preposition would not be possible in this case, as can be seen in (63), compared to use with the preposition to in (64).

(63) *I live from/by/in/on day by day.
(64) ?I live from day to day.

Hence, it is possible to suggest that there is an indication that the N₁ by N₁ construction has developed differently than the N₁ to N₁ construction. It may be the case that the underlying structure N₁ + P + N₁ developed from P₁ + N₁ + P₂ + N₁ with some prepositions and was then used with other prepositions through analogy. Further study of the history of this type of construction is however needed to verify this suggestion.

The next step in lexicalisation described by Lindquist & Levin (2009: 176) might however be true for N₁ by N₁: “When adverbial meaning is specialized [...] hyphens become more frequent”. Albeit not many, there were 87 hyphenated adverbials out of 1361 adverbial instances in the written subcorpus. Specialised meaning was not found at this stage regarding the N₁ by N₁ construction. As exemplified in (65), most hyphenated adverbial instances of N₁
by $N_1$ were very concrete.

According to Lindquist & Levin (2009: 177) “[t]he next step in institutionalization is when the phrase is used adjectivally as a premodifier”. As our corpus data showed, there were many premodifier instances.

(65) Despite government efforts to strike a compromise enabling both mosque and temple to stand side-by-side in Ayodhya, the threat of a communal eruption remains. (A4H, W_newsp_brdsht_nat_report)

The second most common function was the premodifier function. The complementary searches in the OED for the five most common construction types in the BYU-BNC showed that with all types the adverbial function occurred far earlier than the premodifier function. Data also showed that premodifier instances were to a higher extent hyphenated, which, given that the indication that the premodifier function came about later than the adverbial function is true, supports the claim that lexicalisation of items is connected to hyphenation. Lindquist & Levin describe a “fixing of collocates” in that premodifier instances of certain construction types tend to be restricted to certain nouns. This tendency has been identified in the sections on collocations in the present paper. For instance it was shown that case by case collocated most frequently with basis and that step by step often collocated with guide. It was also shown that some types of the construction, when functioning as premodifiers, collocated with certain classes of items. Step by step for instance collocated with items that indicated ‘instruction’ such as guide, tutorial, etc.

According to Bauer (1983: 48), lexicalisation occurs when “the lexeme has, or takes on, a form which it could not have if it had arisen by the application of productive rules”. The final stage in the lexicalisation process, as demonstrated by Lindquist & Levin (2009: 177), occurs when the $N_1$ P $N_1$ construction undergoes nominalisation. There were few nominalised instances of the $N_1$ by $N_1$ construction, but nevertheless they existed. No indications in the diachronic data from the OED could however show if nominalised instances of the construction appeared in newer texts than adverbial or premodifier instances did. It could be argued that when the construction takes on a new form and changes meaning, semantic lexicalisation occurs. A step-by-step, the noun meaning ‘guide’ or ‘instruction’ may have undergone semantic lexicalisation and nominalisation. Further diachronic investigation is
however needed to determine the validity of that possibility.

Data have shown similarities with Lindquist & Levin’s (2009) results. It is possible that the development patterns of instances of the $N_1$ to $N_1$ construction with body-part nouns are similar to those of the $N_1$ by $N_1$ construction. This leads us to the question of whether these constructions are subordinate to the $N_1$ P $N_1$ construction or if they should be regarded as constructions on their own. This issue is discussed in section 5.2.

5.2 The $N_1$ P $N_1$ construction(s)

As mentioned earlier, the pattern $N_1$ by $N_1$ is regarded as a construction in the present study. It was however indicated from the very beginning that the term construction is problematic. Firstly, we have the classic lumping/splitting issue: Is the $N_1$ P $N_1$ pattern one or several constructions? The discussion of our results compared to the results from Lindquist & Levin (2009) above suggests that there are similarities between $N_1$ by $N_1$ and $N_1$ to $N_1$. Obvious differences were however also observed. The first and second stages of development according to Lindquist & Levin could not be indicated with $N_1$ by $N_1$ due to the different semantic properties of by compared to to as well as a lack of diachronic data. It was however suggested that the constituent parts of $N_1$ P $N_1$ may have an effect on the possible development and lexicalisation of the construction(s). On the other hand, there were many similarities between $N_1$ by $N_1$ and $N_1$ to $N_1$. Results from the complementary searches for diachronic data backed the suggestion that the construction was probably used as an adverbial a long time before the premodifier use came about. It was further shown that both $N_1$ by $N_1$ and $N_1$ to $N_1$ could function as adverbials, premodifiers and nouns. Hyphenation was also more common with the premodifier and nominalised forms.

Even though other types of $N_1$ P $N_1$ than those with prepositions to and by are not discussed, it seems reasonable to regard the $N_1$ P $N_1$ pattern as a superordinate construction, and the $N_1$ by $N_1$ construction and $N_1$ P $N_1$ constructions with other prepositions as subordinate constructions. As mentioned, Goldberg (2006: 5) defined constructions as “learned pairings of form with semantic or discourse function” and thus included everything from morphemes to full general phrases in the definition. The so-called unpredictability criterion, that a construction needs to have a meaning or form of its own that cannot be derived compositionally, is both included and excluded from her definition of construction. There is a point in not including unpredictability since, as Goldberg (2006:64) states, patterns can be “stored [holistically in memory] if they are sufficiently frequent, even when they are
fully regular instances of other constructions and thus predictable”. The definition derived from this is however that constructions can be anything from a combination of phonemes to large units. If *construction* refers to any kind of stored pattern, the term becomes less meaningful. As mentioned in the theoretical background (Section 3), Langacker (1987: 82) claims that “there is no fundamental difference distinction between morphological and syntactic constructions”. Since syntax and morphology differ, the view that construction is the same for both morphology and syntax makes the definition of the term even more general. The term would therefore be more useful if the scope is limited. There is however an important point in Langacker’s argument. In some cases a stem and an affix can result in the same or very similar meaning as two individual words. Compare for instance *incompatible* and *not compatible*. To not consider phonemes and morphemes as constructions and stay on the syntactic level however makes the term more useful. Furthermore, affixation is not an issue regarding the N₁ by N₁ construction. It is therefore proposed that a definition of *construction* excludes units below word-level and is applied to the syntactic level. Croft & Cruse (2004: 247) also define construction as a configuration on the syntactic level.

Since it has been shown that N₁ to N₁ and N₁ by N₁ can be considered separate constructions, individual words do effect the construction definition. Given the problem with drawing borders between one construction and another, a hierarchical view is more functional. As Schönefeld (2006: 3) points out, there is also the question of abstractness regarding constructions. A hierarchical view allows for both specified lexical items as well as patterns to be regarded as constructions. Jackendoff (2008:11) also indicates that such a view is favourable in regarding types of the NPN construction as *subconstructions*. An example of a hierarchical view of constructions, using the N₁ by N₁ construction and the N₁ to N₁ construction as examples is shown in Figure 3.

Viewing N₁ P N₁ constructions hierarchically would remove the problem of the lumping/splitting issue. N₁ P N₁ is superordinate to the subordinate construction N₁ by N₁ and *day by day* is subordinate to the superordinate N₁ by N₁ construction. This would solve another problem, whether or not the same combination of words, depending on context, can be regarded as different constructions. The answer is then that they are the same construction, but subordinate constructions to the superordinate *day by day* construction. One construction may also be part of another construction. The from N₁ to N₁ construction would then be subordinate to the the P N₁ P N₁ construction, which includes the N₁ P N₁ construction, which
includes the $N_1$ to $N_1$ construction, and so forth.

Further, semantic properties are taken into consideration. Subordinate constructions to *day by day* can for example be *day by day* (context 1) and *day by day* (context 2).

![Diagram of syntactic and semantic hierarchical view of N P N constructions](image)

**Figure 3. Syntactic and semantic hierarchical view of N P N constructions**

This gives a possible solution to the problem with varying fixedness of the $N_1$ by $N_1$ construction – if the semantic properties of the construction differ between e.g. hyphenated (which may be argued to be influencing fixedness) and unhyphenated instances, it seems wrong to regard the hyphenated and unhyphenated instances of the construction as the same construction, but intuition still tells us that they are. There is also the question of whether or not the pattern can be regarded as one construction or several depending on its position (i.e. syntactic function) within sentences. This is also solved by viewing the construction as part of a hierarchy. Similar to determining individual subordinate constructions due to different contexts, instances of the construction with different functions can be regarded as individual subordinate constructions. Further, there is the possible semantic influence of the individual words and morphemes within the construction and how this affects the classification. As mentioned in the theoretical background, language processing and how lexical items are

---

22 It should be noted that only subordinate constructions that are relevant to the results from the present investigation are mentioned here. NPN may be superordinate to other patterns as well.
retrieved from memory varies.

There are various ways of describing the retrieval of information from memory. Sinclair (1991: 109–115) presented two possible means of language production: the idiom principle and the open choice principle. The first one means retrieving semi-preconstructed phrases from memory and the second one applying semantic information to a structure, i.e. filling grammatical slots with words. By viewing constructions hierarchically, two identical phrases can be produced using different principles but still constitute the same construction and at the same time be two different subordinate constructions. Jackendoff (2008: 18) touches on this problem in his discussion of the N P N construction. He also mentions the problem of whether the construction is “in the grammar” or “in the lexicon” and proposes that the nature of the lexicon be reconsidered, that “[i]t is not a repository of words: it is a repository of whatever pieces of linguistic structure have to be stored in long-term memory” (ibid.). In that case it is plausible that different meanings of the construction are also stored together with it in the lexicon. This leads us to a suggestion about semantic issues regarding the construction. If one instance of the pattern indicates succession and another indicates time – is it still the same construction? The answer is then that they are different subordinate constructions semantically to the superordinate construction which is one construction syntactically. The same construction might even be stored twice in the lexicon with different semantic information connected to it.

To further exemplify the benefits of this view we take the on a step by step basis construction. Figure 4 shows it in the construction hierarchy of on a N P N basis.

In the proposed model, this construction would subordinate to the on a N1 by N1 basis construction, which would include the N1 by N1 construction and be subordinate to the on a N1 P N1 basis construction. Since it is possible to use other prepositions than by, another subordinate construction to the on a N1 P N1 basis construction would be on a N1 to N1 basis, as exemplified by the subordinate construction on a day to day basis.
6. Summary and conclusions

In the present study, the \( N_1 \) by \( N_1 \) construction was investigated in terms of function (in sentences), frequency, fixedness (and lexicalisation) and semantics. The first research question was:

- What are the significant characteristics of the \( N_1 \) by \( N_1 \) construction, distributionally, syntactically and semantically by function, form\(^{23}\), collocation patterns, mode\(^{24}\) and text-type?

Distributionally it was concluded that the construction occurs frequently in British English. In the BNC it had a frequency of 20.5 instances per million words, based on types that occurred more than once. In this respect there was a difference between written and spoken language: the construction is less frequent in the spoken mode. Also, there were differences between the written text-types: the frequency per million words was lower in the \textit{academic} and \textit{newspaper} subcorpora. It was suggested that this is due to reasons of formality. Conclusions about syntax and semantics could also be drawn from the distributional data. Regarding function, the

\(^{23}\) Unhyphenated or hyphenated.

\(^{24}\) Spoken or written English. The fifth subcorpus contains the transcriptions of spoken language.
adverbial function was the most frequent one, far more frequent than the premodifier function, and thereby considered the standard form. It was also shown that the frequency of the forms – hyphenated and unhyphenated – corresponded with functions. The adverbial function scarcely occurred hyphenated, but the premodifier function was hyphenated in a majority of cases. No major differences regarding function could be identified when comparing the written and spoken subcorpora. For understandable reasons, hyphenation was not identified in the spoken mode. Regarding text-type there were differences in the proportions between adverbials and premodifiers. The fiction subcorpus stood out in that premodifiers were much rarer than in the other written subcorpora and consequently hyphenation as well. Other functions than adverbials and premodifiers were few, and most of them were instances that could not be classified since they constituted headlines or similar. A few instances of the construction were nouns and these were also hyphenated. It needs to be pointed out again that these results are based on types of the construction that appeared more than once in a subcorpus. Consequently, the results could be somewhat different if types that only appeared once were included. Albeit it is very unlikely that results would avert from the general tendencies that were demonstrated in the present paper.

A collocation analysis was carried out for some types of the construction that showed specific tendencies. The most important finding was the high frequency of the type case by case collocating with basis in the on a/an X N₁ by N₁ basis construction. It was also found that step by step collocated frequently with nouns that indicated ‘instruction’ such as guide and approach.

The second research question was:

- Are there signs of lexicalisation with this construction?

It was suggested that the N₁ by N₁ construction might be undergoing a process of lexicalisation. Like Lindquist & Levin (2009) have demonstrated for the N₁ to N₁ construction with body-part nouns, diachronic data indicated that the N₁ by N₁ construction may follow a pattern of development where it started off as an adverbial and later acquired premodifier use. Nominalised and hyphenated forms were also present in the synchronic data. However, to show definitely whether or not this is a process of language development requires diachronic studies of the construction. Based on the present material it is only possible to draw
synchronic conclusions and show some diachronic indications, but a reasonable assumption based on the diachronic data and synchronic conclusions is that development started with the adverbial function. These results are similar to those found by Lindquist & Levin (2009) which makes N₁ P N₁ constructions with other node prepositions interesting for further study.

There was also a theoretically driven research question:

- Is the notion of construction useful in the study of N₁ P N₁ and is N₁ by N₁ a type of a construction, a construction of its own or several constructions?

The notion of construction is useful in the study of N₁ P N₁ if a clear definition of it is provided. To regard everything from phoneme to sentence as constructions renders the term useless, but limitations such as staying on the syntactic level and including semantics in the definition of the term increases the usefulness. In this study a hierarchical view of the N₁ P N₁ construction has been suggested. This is not intended as a final definition of the term construction but rather a suggested angle of incidence when forming a workable definition of construction. Regarding the definition of N₁ by N₁ it has in this view been decided that to view the N₁ by N₁ construction as one superordinate construction with many subordinate constructions that share the same syntactic pattern.

Construction grammar has differentiated itself from other grammatical theories in its focus on constructions, but in order for the framework to really be useful for the study of linguistic phenomena, the discussion around what a construction is needs to be intensified and a common construction grammar oriented definition of construction needs to be adopted.

Finally, regarding corpus linguistic methodology and the present topic of study it can be concluded that the investigation of the N₁ by N₁ construction using the corpus linguistic method has yielded several important results that could not have been achieved through approaches based on intuitive or qualitative study rather than quantitative results. In this respect, corpus linguistics is an excellent means of investigating general syntactic and lexical tendencies. It can also show variation through investigation of frequencies, and most importantly, especially regarding written language, it shows the actual use of language by speakers of the language, hence reducing the number of conclusions mainly based on intuition. In combination with qualitative semantic analysis, the quantitative data corpora give
us helps to increase our understanding of language in specific contexts, such as understanding the N₁ by N₁ construction, as well as language in general.
References
Primary sources


Secondary sources


