Like I said again and again and over and over:

On the ADV1 and ADV1 construction in English¹

Magnus Levin, Linnaeus University
Hans Lindquist, Malmö University

¹ We would like to dedicate this paper to Michael Stubbs, who with his theoretical astuteness and methodological ingenuity has served as source of inspiration for our work over many years, during which he has also been a supportive friend.
1. Introduction: The ADV1 and ADV1 construction

The role of multi-word units in language and their relevance for linguistic theory has attracted increasing interest over the past three decades (cf., e.g., Pawley & Syder 1983; Altenberg 1998; Moon 1998; Sinclair 1999; Wray 2002; Stubbs 2002; Ellis 2008; Granger & Meunier 2008; Gries 2008; Biber 2009). Corpus methodology has made new types of empirical investigations possible and has moved the focus from infrequent idiomatic expressions to frequently recurring word combinations. A large number of labels have been proposed to denote such multi-word units of more or less overlapping types (see Wray 2002: 8 – 10 for a comprehensive list); in the present study we will follow Wray and use the term ‘formulaic sequence’ and, occasionally, the term ‘construction’ in a non-construction-grammar sense.

Stubbs (2002) discusses two types of multi-word units: collocations of content words and “phrases which consist of a combination of grammatical and content words” (p. 227). The latter can be described by formulae like “PREP the NOUN of the” (ibid. 232) and illustrated by examples like at the end of the. In a later paper, Stubbs distinguishes between “phrase frames”, which are n-grams with a variable slot like plays a * part, and “PoS-grams”, which are strings of part of speech categories like preposition + determiner + singular noun + preposition + determiner (2007a: 90 – 91). The units studied in the present paper, sequences like over and over, do not fit comfortably into any of these categories since they consist of only grammatical words. For the same reason, they are neither collocations nor colligations in the terminology of Sinclair (1998). If anything, they can be said to be a special kind of PoS-gram with the form ADV1 and ADV1, i.e. a combination of two identical adverbs coordinated by the conjunction and. They belong to a type of sequences consisting of symmetrical structures with two items of the same word class joined by some linking word (cf. Lindquist & Levin 2009: 173 – 174). There seems to be a tendency in English and many other languages to favour such sequences. In an article on the PoS-gram “noun-preposition-noun” (in his construction grammar terminology: the NPN construction), Jackendoff (2008: 15) suggests that such constructions are “cognitively natural”. In the same article, he mentions a “family of idioms”: back and forth, up and down, to and fro and round and round.

---

2 Biber (2009: 275) makes the same distinction between two types of multi-word sequences: “‘multi-word lexical collocations’ (combinations of content words)” and “‘multi-word formulaic sequences’ (incorporating both function words and content words)” which he sees as the two poles of a cline (ibid. 290).
which contain combinations of either identical or different adverbs, joined by the conjunction
*and* (ibid. p 12). The present paper focuses on combinations with identical adverbs, i.e. the
ADV1 *and* ADV1 construction, disregarding for the time being sequences of different
adverbs like *back and forth* etc.

The ADV1 *and* ADV1 construction is theoretically interesting for several reasons. As
pointed out by Stubbs (2007b:165) about phrases in general, “[t]he semantic denotation of the
individual word is weakened and the pragmatic meaning of the multiword-sequence is
strengthened”. The sequences under study illustrate a number of typical paths of such
semantic/pragmatic change from concrete to abstract – textual or subjective – meanings (cf.
Brinton & Traugott 2005; Traugott 1995; Traugott & Dasher 2005). Through the use of
historical corpora, we are able to study these diachronic processes and establish paths of
change for a number of instantiations of the construction. We can also study the relation
between frequency and linguistic change (cf. Barlow 2011:24–25). Furthermore, our findings
shed some light on the interplay between speakers’ application of grammatical rules and their
use of ready-made phrases, and thus on the question of what constitutes the units of the
language (cf. Stubbs 2007a:90), and what is the role of the lexicon (cf. Wray 2002:261–265).
Finally, we are able to study the influence of register on various developments. In particular,
we are looking for answers to the following questions:

- What syntactic roles do the sequences fulfil and how have these roles developed
  over time?
- How have the frequencies of the most frequent ADV1 *and* ADV1 sequences in
  English changed over time?
- What kind of meanings do they express and how have these meanings developed
  over time?
- How are these changes related to differences between formal/informal and
  written/spoken registers?
- To what extent can these ADV1 and ADV1 sequences be said to form a set or a
  “family” of constructions which have similar form and meaning and display similar
developments?

2. Material and method
Our main data sources are The Corpus of Historical American English (COHA) (Davies 2010–) and The Corpus of Contemporary American English (COCA) (Davies 2008–); for discussions and explorations of COCA, see Davies (2009, 2010). In addition, we make some comparisons with The British National Corpus (BNC), The Longman Spoken American Corpus (LSAC) and the new 155-billion-word Google Book Corpus covering 1810–2009 (Davies 2011–).

COHA contains 400 million words from the years 1810–2009. It enables researchers to carry out diachronic studies of rare phenomena, but due to its composition the corpus has some limitations. It contains four different genres, Fiction, Magazines, Newspapers and Non-Fiction. The composition of the sub-corpora is not, however, fully consistent over time. The fiction sub-corpus mainly contains novels (from Moby Dick to Hollywood Wives), short stories, and drama, but also, for the 19th century, poetry, and, for the 20th century, movie scripts (from Singin’ in the Rain to Psycho). The magazine sub-corpus contains a wide variety of sources from Harpers, The Nation, Sports Illustrated to Time Magazine. About half the newspaper material comes from The New York Times, and an additional 10% each from The Chicago Tribune and The Christian Science Monitor. This focus on rather few newspapers means that findings from the newspaper sub-corpus must be treated with some caution. The biggest categories of non-fiction books include history, social sciences, science and technology.

The genre distribution in the corpus varies over time, which makes diachronic comparisons slightly problematic. The proportion of fiction texts remains around 50%, while the coverage of newspapers and non-fiction is more restricted. Newspapers only occur in COHA from the 1860s and only constitute around 7% of the corpus size per decade until the 1920s. Thus, overall findings from the corpus are largely based on the language of fiction. Comparisons across the whole span of the corpus are also affected by the fact that the 1810s sub-corpus only contains around one million words. Some of our comparisons over time therefore had to be restricted to the 1850s onwards.

COCA is continuously updated and in the version we used contained 425 million words from 1990–2011. This corpus comprises roughly equal proportions of Spoken, Fiction, Magazines, Newspapers and Academic journals. The spoken sub-corpus contains material from news broadcasts and talk shows, such as Good Morning America, 60 Minutes and Jerry Springer. Because there is no spontaneous conversation (see, however, under Texts/types at http://corpus.byu.edu/coca/ for a discussion of the naturalness and authenticity of the spoken
data), we also compare with LSAC. This is a corpus of spontaneous conversations from the 1990s of about 5 million words (Leech et al. 2009:xxi, 100). The Fiction sub-corpus in COCA consists of short stories from literary magazines, first chapters of novels, and movie scripts. There are about 100 different titles in the magazines sub-corpus, ranging from *Cosmopolitan* and *Christian Century* to *Sports Illustrated*. The newspaper sub-corpus is sampled from ten different publications, such as *The New York Times* and *The San Francisco Chronicle*. Finally, the Academic sub-corpus contains about 100 different peer-reviewed journals, such as *American Scholar* and *Journal of Environmental Health*.

The BNC (Aston & Burnard 1998) was used to compare the American material with British English from the 1990s. For this we utilized Mark Davies’ interface (http://corpus.byu.edu/bnc/).

As will be seen in the next section, we began our survey of the ADV1 and ADV1 construction by searching COCA and the BNC for sequences with two adverbs coordinated by *and*. We then searched for the most frequent sequences in the COHA material limiting our investigation to five different decades, the 1810s, 1850s, 1900s, 1950s and 2000s.

### 3. Results

#### 3.1 Distribution

##### 3.1.1 Overall frequency

Table 1 presents an overview of the twelve most frequent sequences of identical adverbs of direction or position in the COCA. Frequencies for the BNC are supplied as a comparison. Sequences of two identical comparatives like *more and more* (2870 tokens per 100 million words in COCA), *less and less* (315), *faster and faster* (109), *further and further* (81) etc. have been left out. Similarly, *so and so* has been left out since its function is very different from that of the others. The inclusion of *again and again* could perhaps be questioned since in Present-day English *again* normally expresses neither direction nor position, but it is motivated by the fact that the original meaning of *again* was “[i]n the opposite direction; back” (*OED* s.v. again A. adv. 1.a.), and an even stronger motivation is that *again and again* functions very much like *over and over*, as will be seen below.

**Table 1.** The twelve most frequent sequences of identical adverbs of direction or position in COCA and the BNC. Sequences of two identical comparatives and *so and so* excluded. Frequency per 100 million words.
<table>
<thead>
<tr>
<th>Rank in COCA</th>
<th>COCA per 100 million words</th>
<th>Phrase</th>
<th>BNC per 100 million words</th>
<th>Rank in the BNC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1719</td>
<td>over and over</td>
<td>540</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>1027</td>
<td>again and again</td>
<td>649</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>699</td>
<td>on and on</td>
<td>395</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>122</td>
<td>around and around</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>96</td>
<td>round and round</td>
<td>109</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>80</td>
<td>up and up</td>
<td>75</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>60</td>
<td>through and through</td>
<td>63</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>29</td>
<td>by and by</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>9</td>
<td>28</td>
<td>down and down</td>
<td>38</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>22</td>
<td>out and out</td>
<td>31</td>
<td>8</td>
</tr>
<tr>
<td>11</td>
<td>16</td>
<td>back and back</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>in and in</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>3900</td>
<td></td>
<td>1935</td>
<td></td>
</tr>
</tbody>
</table>

Raw figures. Instances of triplication like *over and over* and *over and over* are here counted twice and instances of hesitation in spoken language have not been disregarded. However, instances where the sequence is not a coordination of two adverbs, like (*…*) before we did that we had to put the natural plug in and in *that plug there were five like aluminium coloured er pins* (*…*), have been disregarded.

The frequencies in Table 1 can be compared with the cut-off frequency used by Biber *et al.* (1999:990) in their studies of so-called lexical bundles (i.e. any recurring word strings), which was 10 occurrences per one million words (i.e. 1,000 occurrences per 100 million words).

With that criterion, only *over and over* and *again and again* in the American corpus would qualify. Nevertheless, it is possible to discern a set or family of similar ADV1 and ADV1 sequences where the top ones have a fairly high frequency and the rest are related by having the same form. Except for minor variations, the top-ranking sequences are the same in American and British English. The most striking difference is that the un-hyphenated sequence *by and by* only occurs twice in the BNC (spelt *bye and bye*), together with two instances with hyphens. Hyphenated instances are not included in the table, but for some of these sequences there are a fair number of hyphenated tokens in COCA: *up-and-up* 7 per 100 million words, *through-and-through* (3), *out-and-out* (10), *by-and-by* (3). We will return to the question of hyphenation below.
The difference in frequency between *around and around* and *round and round* in the two varieties can be explained by the well-known regional differences between the single words *around* and *round* (see, however, Section 3.5 below for a modification of this view). Table 1 also shows that three of the sequences are considerably more frequent in American English than in British English: *over and over*, and, to a lesser extent, *again and again* and *on and on*. We will return to this fact below. As will also be shown below, over time many of these sequences have undergone semantic developments and have acquired specialized and partially overlapping meanings, which in turn may have influenced frequency.

In the remainder of the paper, we will focus on the most frequent sequences and those that show the most interesting developments.

3.1.2 Frequency over time

The frequency of the ADV1 and ADV1 sequences has changed notably over time. The developments for a number of them are illustrated in Figure 1.

![Figure 1. The distribution of again and again, on and on, over and over, by and by, (a)round and (a)round, through and through, up and up, out and out and down and down in COHA, per million words](image)

3 Unless otherwise specified, all numbers in figures and tables have been post-edited to exclude irrelevant tokens. 'Multiple' tokens (*on and on and on*) have been counted as one instance each. Hyphenated instances are
As is shown in Figure 1, three of the sequences increased significantly during the period: *over and over, again and again* (peaking in the 1900s) and *on and on*. *Around and around* also rose, but at a much lower level. At the same time, several sequences decreased in frequency: *round and round* (after having peaked in the 1850s) and *through and through* (after having peaked in the 1900s). Perhaps the most striking development, however, is seen with *by and by*. This sequence is quite frequent in the mid- (to late) 19th century, but already in the early 1900s it had started a steep decline and by the mid-20th century it had virtually disappeared from use.\(^4\)

These results do not appear to be based on random fluctuations in COHA, since similar changes over time were found in the Google Book Corpus: *Over and over, on and on* and *around and around* have been slowly increasing during the last two centuries, while *again and again, by and by, round and round* and *through and through* all peak around 1900 and then decrease. The Google Book Corpus can thus serve to support or disconfirm hypotheses from other corpora.

LSAC provides some additional support for these developments. *Again and again*, which appears to have been decreasing for a century, only occurs 1.2 times pmw (6 tokens). *On and on* and *over and over*, which both are still on the increase in COHA, are more frequent than *again and again*, but *over and over* is rarer in LSAC (9.8 pmw; 49 tokens) than in COHA (17.5 pmw) and *on and on* is on a roughly similar level, 7.8 pmw (39 tokens) in LSAC as compared to 7.2 pmw in COHA.\(^5\)

We thus see that the individual sequences have individual histories, and beginning with section 3.2 below the most interesting will be dealt with one by one. We will then return to a discussion of more generalized aspects in the Discussion section.

---

\(^4\) The decrease of *by and by* coincides with frequencies for other adverbials meaning ‘soon’ (*before long, presently, soon*) in COHA decreasing by 50%. Instead, some adverbials meaning ‘in the end’ (*eventually, finally, in the end*) have more than trebled in frequency. The demise of *by and by* therefore seems partly explicable in terms of a general change in the use of temporal adverbials in English. The reasons behind this shift require further study.

\(^5\) LSAC also supports the findings for the other adverbs: *Round and round* and *around and around* are roughly equally frequent in LSAC (in spite of the single *around* being many times more frequent than *round* in that corpus, as in most corpora). *Through and through* (3 tokens), *by and by* (0 tokens), *out and out* (2) and *up and up* (3) were, as could be expected, very rare in the spoken corpus. Comparisons between LSAC and the Spoken sub-corpus of COCA show that the less spontaneously produced speech from COCA has consistently higher frequencies for all three sequences: *again and again* (10.5 pmw), *on and on* (12.5 pmw) and *over and over* (25.0 pmw).
3.2 Over and over

One of the definitions of over given by the OED (s.v. over III.7b) is: “So as to make what was an upper surface a lower one, by turning it forward (or laterally) and downwards, or to turn something upside down. Also reduplicated to indicate repetition of such motion, as over and over.” The OED also notes the extended, metaphorical meaning (s.v. over III.15): “repeatedly, many times over. Also over and over again.” In our material there is a difference between over and over and over and over again in that motion verbs do not occur with over and over again, cf. (1) and (1a).

(1) There were three large coins and one smaller one. He turned them over and over, and finally ascertained that the large coins were ten dollar pieces, and the smaller one a five dollar piece. (COHA; Fiction; 1856)

(1a) *he turned them over and over again

In contrast with again in the sequence again and again (see below), the adverb over retains a concrete meaning component denoting motion or change of position in space even when the sequence occurs with the meaning ‘again’. The connection with physical motion, however, has weakened over time. This is illustrated in Figure 2, which shows the proportion of motion verbs used with over and over in COHA over the last 150 years.

![Figure 2. The percentage of motion verbs with over and over in COHA](image-url)
As Figure 2 shows, motion verbs, in this case almost exclusively *turn* and *roll*, decrease steadily in frequency over the period.\(^6\)

In a parallel development, the use of *again* after *over and over* as in (2) has decreased, so that the most common usage is now *over and over* without *again*, as in (3).\(^7\)

(2) “My son -- my dearest son,” she said, *over and over again*, […]. (COHA; Fiction; 1859)

(3) She said Goddammit, goddammit, goddammit, *over and over*, like that. (COHA; Fiction; 2000)

This trend is illustrated in Figure 3.

![Figure 3. The percentage of over and over followed by again in COHA. Motion verbs excluded.](image)

There is no indication that *again* is lost faster in collocations with low-frequency verbs than in those with high-frequency ones (there was instead a non-significant trend suggesting

---

\(^6\) The differences between the 1900s, 1950s and 2000s are statistically significant (chi-square test, \(p \geq 0.05\)). The figure is based on the following numbers: 1850s 20.3\% (28 motion verbs in 138 tokens), 1900s 13.8\% (44/317), 1950s 8.2\% (27/330) and 2000s 4.6\% (24/518).

\(^7\) One of our reviewers has pointed out that this is reminiscent of Jespersen’s cycle of sentence negation, which indeed it is in the sense that the adverb originally carrying the meaning, *again*, disappears and only the intensifying modifier, *over and over*, remains.

\(^8\) All changes in Figure 6 are significant (\(p \geq 0.05\)), except between the 1900s and the 1950s. The figure is based on the following numbers: 1850s 79.1\% (87/110), 1900s 44.0\% (120/273), 1950s 40.6\% (123/303) and 2000s 29.6\% (146/494).
that *again* is lost slightly more rapidly with high-frequency verbs). Bybee (2006:728–729) argues that high frequency constructions (such as *no* negation) are resistant to change. However, in the present case the change involves a reduction in form, and it seems likely that the reduction can proceed faster in contexts where readers readily recognize the collocation (as is the case for phonological reduction, see Bybee 2006:714–715).

American English is leading the way in the dropping of *again* after *over and over*. The distribution over genres in COHA can be compared with that in the more recent corpora COCA and BNC presented in Figure 4.

![Figure 4. Over and over followed by *again* in the different genres in COCA and the BNC](image)

Perhaps contrary to expectation, the highest proportion of *again* following *over and over* is found in the Spoken and Academic sub-corpora in COCA and the BNC, while the lowest is in Fiction. Furthermore, in spontaneous AmE conversation from LSAC, 49% (24/49) of the instances of *over and over* are followed by *again*, which suggests that the loss of *again* is not an effect of increasing colloquialization. One possible explanation for the retention of the longer form in the spoken language is that wordiness is more acceptable there than in written genres, where more succinctness is expected. The COCA and BNC figures are still puzzling,

9 The top five verbs (*SAY, REPEAT, TELL, DO, PLAY*) retained *again* in 22% of the instances (32/145) in the 2000s as compared to 32% (114/351) with the less frequent verbs.

10 It was not possible to go through the nearly 8,000 tokens manually to identify motion verbs (which do not permit variation with *again*) and multiple instances (*over and over and over (again*)). Figure 8 is therefore based on raw numbers.
however, since it is usually Fiction that is most spoken-like, and the high proportion of again in Academic is surprising.

To conclude, in approximately 60% of the cases in COCA, the full meaning of ‘repetition’ is expressed by over and over without the added again. In the BNC, on the other hand, 61% (337/555) of the cases retain again. This suggests that over and over is becoming synonymous with again and again in American English; it may even be on its way to take over the earlier role of again and again. This hypothesis is supported by the frequency figures in Table 1 above, where it was shown that while again and again is still the most frequent sequence in British English, over and over is now the most frequent one in American English.\(^{11}\)

3.3 Again and again

In the sequence again and again, a word denoting ‘repetition’ is itself repeated. The OED (s.v. again A.4a) gives the meaning of again as “Repetition of an action or fact: another time; once more; any more; anew; […]”, and notes that “[t]his sense is more fully expressed by once again, over again; and the repetition increased by too and again (obs.), again and again, ever and again, time and again.” (ibid. 4b).\(^{12}\)

In order to get at the meaning and possible changes in meaning of the adverb sequences over and over and again and again we investigated with which verbs they were used and how freely they were used with different verbs. The similarity in meaning between over and over and again and again is indicated by the fact that they are used with very similar verb types. Diachronically, the type of verbs occurring most frequently with both adverbs has remained stable. In the 1900s, 6 of the 11\(^{13}\) top verbs with again and again also occurred with over and over (KISS, REPEAT, GO, SAY, READ, TURN), and the same was the case in the 1950s (although for partially different verbs: SAY, HEAR, REPEAT, TELL, GO, ASK). In the 2000s, the proportion was 7/11 (TELL, REPEAT, SAY, DO, ASK, GO, READ).

---

\(^{11}\) Another, less frequent, alternative to again and again and over and over (again), without repetition of any adverb, is over again: I said that during the campaign. And I’ve said it over again as governor (COHA; Spoken; 2011). Usually, (all) over again has the slightly different meaning “anew”: It was like learning to fly all over again (COCA; Fiction; 2011).

\(^{12}\) Of the alternatives mentioned by the OED, too and again is indeed obsolete, only occurring in one text from 1923 in COHA, ever and again only occurs between 1847 and 1947 and time and again is infrequent and decreasing slightly in the 20th century.

\(^{13}\) In the 1900s, four verbs were in joint eighth position in the order of frequency, and in the 2000s, two verbs shared tenth place. Hence the inclusion of the top eleven collocates.
An analysis of the sequences in context indicates small differences in preferences between the adverbs and a high degree of interchangeability, as suggested by (4) and (5).

(4) Mother had told me again and again not to be angry with him; (COHA; Fiction; 2000)

(5) I must not feel guilty, Mama and Papa told me over and over, in the days afterward, […] (COHA; Fiction; 2006)

The COHA material suggests one minor, but significant, difference between the two adverbs: negative connotations such as dislike, annoyance, or sarcasm as in (6) are significantly ($p \geq 0.05$) more often expressed with over and over (50% (260/518)) than with again and again (41% (151/369)).

(6) Oh, Mark. It's the same old thing, over and over. We need something new. (COHA; Fiction; 2006)

Even with motion verbs like TURN, the adverbs are used with similar meaning, as in (7).

(7) Dotty groaned in her sleep and turned again and again. (COCA; Fiction; 2000)

There is little difference between again and again and over and over when it comes to collocating verbs, and they are both most common in fiction. The differences found in COHA are instead restricted to proportions in different genres – again and again being more than twice as frequent in fiction than any other genre while over and over is only slightly more common in fiction – and to diachronic developments: over and over appears to be slowly replacing again and again.

3.4 On and on

One of the meanings given for on in the OED (s.v. on A.4.a.) is: “Onward or forward in space, time, or condition. […] Sometimes repeated for emphasis, as on and on […]”. For on

---

14 In the present study negative attitudes were either deduced from the immediate context such as the co-occurrence of a verb expressing negative connotations (She looked almost like an angel, that is if angels had braided hair and prattled on and on. (COHA; Fiction; 2009)), but often wider contexts such as the sentence or paragraph needed to be taken into account.
and on, the OED gives (s.v. on A.5.b.): “at length; again and again, continually, eternally; now freq. depreciatively, esp. with verbs indicating speech: interminably, incessantly.” The OED thus acknowledges that there has been a semantic development from denoting position or motion in physical space to denoting extension in time and even repetition (although on and on normally refers to durative action and again and again to punctual repetition), and furthermore notes that the sequence has acquired negative connotations. In this process, the number of different verbs used with on and on has changed, so that although the number of tokens increased steadily during the period, the number of types reached a peak with 56 in the 1900s and then decreased to 37 in the 2000s. During the same period, tokens with motion verbs decreased and tokens with the verb GO increased.

Typical motion verbs in earlier usage were, e.g., GO, RIDE, and WALK, as in (8). In the analysis, GO was classified as a motion verb in those cases where physical motion was evident (common until 1900s), as in (9).

(8) As he rode on and on remorse drew him into his grasp. (COHA; Fiction; 1909)
(9) On and on they went, making slow progress because the trail was very poor. (COHA; Fiction; 1909)

The decrease in motion verbs and the increase in the verb GO is made evident in Figure 5.

Figure 5. The percentage of motion verbs and the verb GO with on and on in COHA

The numbers for GO in the four decades were 2, 30, 87 and 120, and the numbers for motion verbs were 12, 79, 23 and 17.
The increase in the use of *go* fits the typical S-curve often seen in language change, while the motion verbs follow the corresponding downwards trend.

In conjunction with the decreasing number of motion verbs and increasing use of the verb *go*, there has been an increase in negative connotations expressed with *on and on*. This is illustrated in Figure 6, where the differences between the 1900s, 1950s and 2000s are significant ($p \geq 0.05$).

![Figure 6. Negative connotations with *on and on* in COHA](image)

In the 2000s, *go on and on* either refers to negative events continuing incessantly, as in (10), or to speech events that, according to the speaker, continue for too long as in (11).

(10) When he fell ill, the suffering went *on and on*. (COHA; Fiction; 2001)

(11) […], once Marv gets going about his car, it’s downright pain-in-the-arse material. He *goes on and on*, like a kid, and he’s just turned twenty, for Jesus’ sake. (COHA; Fiction; 2006)

Apart from *go*, *on and on* also collocates with some (often negative) verbs of communication, e.g. *drone, talk* and *prattle*.

*On and on* differs from *again and again* and *over and over* in that it normally occurs with verbs construed as durative as in (12), while the latter occur with verbs construed as punctual, as in (13) and (14).

---

16 The numbers for negative connotations were the following: 1850s 15.8% (3/19), 1900s 18.4% (25/136), 1950s 54.2% (90/166) and 2000s 76.2% (163/214).
(12) [...] the Salvation Army band could be heard still playing cheerfully *on and on*. (COHA; Fiction; 1953)

(13) Basically, Brendan’s style is to play that Missy Elliott album *again and again* until I want to kill him. (COHA; Magazine; 2003)

(14) Sometimes even horrible memories play *over and over* in my mind, [...]. (COHA; Fiction; 2003)

In conclusion, *on and on* has developed two new kinds of meaning: durative and textual. In addition it has gone through subjectification, expressing mainly negative attitudes on the part of the speaker.

### 3.5 *(A)round and *(a)round

As mentioned in the introduction, the adverb forms *around* and *round* are variants in more or less free variation. The same seems to be the case for the sequences *round and round* and *around and around*, which will therefore be treated together here. According to many sources, *around* is generally preferred in American English and *round* in British English (cf., e.g., *OED* s.v. *round adv.* and *prep.*). However, Baker (2011:77–80) has shown that in British English as represented by the Brown family of corpora, the total frequency of *around* overtook *round* between 1961 and 1991, and then increased very fast between 1991 and 2006. It is noteworthy, however, that the sequence *round and round* is still much more frequent than *around and around* in British English.

The preference in American English for *around and around* over *round and round* is also quite recent: *round and round* was the most frequent form in COHA all the way up to the 1950s. It is only in the 2000s that *around and around* has prevailed in the American corpus. This small current preference for *around and around* should be compared with the overall frequencies of the single adverbs *around* and *round* in COHA, where *around* is 25 times more frequent than *round* in the 2000s. Adverbial *round* has thus been preserved to a greater extent in the ADV1 and ADV1 construction than in the language as a whole both in British and American English, which suggests that the older form of the adverb has been “fossilized” inside the construction in the same way as obsolete words may live on in formulaic sequences, as mentioned by Wray (2002:49).
One of the meanings given by the OED for round is (s.v. round A.I.3.b.): “In a circular or orbital course; so as to make a complete circuit. Freq. reduplicated to indicate repetition, as round and round.” This sense is illustrated in (15).

(15) Jack could hear them muttering as they turned the strange black objects around and around in their hands. (COHA; Fiction; 2004)

Our findings thus show that although the simple adverb around is increasing rapidly at the expense of round, this process goes considerably slower in the sequence (a)round and (a)round. The reason for such fossilization, Wray (2002:130) argues, is that native speaker children engage in Needs Only Analysis in their acquisition process, meaning that they analyse and segment linguistic material only if there is a good reason for it (rather than segmenting as much material as possible).

3.6 Up and up

Up and up is usually used with non-figurative meaning, expressing continuous movement upwards as in (16). This is the most frequent use of up and up.

(16) The baby turtle uses her flippers to climb up and up. (COCA; Magazine; 2010)

As noted in the OED (s.v. up n. 4a and 4b), up and up has also developed two metaphorical meanings where up has been converted into a noun, as in (17) and (18). Of the 28 instances of up and up in the 2000s in COHA, three occurred in this idiomatic meaning.

(17) Are you confident that the Colombian officials you dealt with are on the up-and-up and you can trust them? (COCA; Spoken; 1990)

(18) Each article or book chapter counts for one point; each book counts for ten; and the productivity trend, needless to say, is on the up and up. (BNC; Non-Fiction; 1990)

In (17), on the up-and-up means ‘honest’, while in (18) it means ‘rising’. In our present-day American material, only the ‘honest’ meaning occurs, while both meanings can be found in the BNC. Hyphenation is variable.
3.7 Through and through

In its treatment of *through and through*, the *OED* gives the following explanation (s.v. through *prep.* and *adv.* A.1.h.): “repeatedly through; so as to penetrate both sides or surfaces of; right through, entirely through. Also *fig.* (Cf. B. 5.)”. It goes on: “With repeated or complete penetration; through the whole thickness or substance; completely from beginning to end; right through, entirely through” (ibid. B.5.a.). And further: “In all points or respects; thoroughly, wholly, entirely, out and out” (ibid. B.5.b.).

With the sequence *through and through* there has been a shift from verbs expressing physical motion or penetration (*PIERCe*, *SHOOT*, *PENETRATE*), as in (19), to the BE ADJ/NP *through and through* construction expressing either positive or negative attitude, as shown in (20). In the COHA material there is a significant shift ($p \geq 0.05$) from physical motion use to abstract extension between the 1850s (93% (37/40) motion verbs) and the 1900s (34% (20/59) motion verbs).

(19) I’ll mount my horse and ride both day and night Till I find out the brute, and then this sword Shall stab him *through and through*. (COHA; Fiction; 1815)

(20) He’s poisoned the child with his crazy talk. Anti-American *through and through*. (COHA; Fiction; 1958)

Syntactically, *through and through* can either postmodify an adjective, as in (20), or an NP, as in (21).

(21) […] that guy is not only rich and spoiled, he’s a bad apple *through and through*. (COHA; Fiction; 2003)

In Present-day English this sequence typically expresses opinions about another person’s character. It thus instantiates a shift from concrete to abstract meaning combined with speaker attitude towards the degree or extent of a particular phenomenon.

3.8 By and by

*By and by* is the only sequence in our material that only occurs in a figurative sense, described by the *OED* (s.v. by and by *adv.* and *noun* A. *adv.* 4) as: “Before long, presently, soon,
shortly.” The original, now obsolete meaning was, however, the non-temporal “Of a succession of (persons or things): One by one, one after another, in order” (OED s.v. by and by adv. and noun A. adv. 1). An example of the current temporal meaning is given in (22).

(22) We’ll understand it all by and by. (COCA; Magazines; 2011)

As noted in the OED, there is also a nominalized form of by and by, ‘time coming’, now mainly used in the form the sweet by and by17 as in (23).

(23) No, you know – God will take care of the sweet by and by. (COCA; 2006; Spoken)

As mentioned in Section 3.1.2, by and by has almost totally gone out of use.

3.9 Out and out

In addition to relatively rarely expressing continued concrete or abstract outward movement, as in (24), the sequence out and out has developed the figurative meanings ‘thoroughly, completely, entirely; downright’ as an adverb, and ‘complete, thoroughgoing, unqualified’ as an adjective (OED s.v. out and out). These latter and more frequent uses are illustrated in (25) and (26).

(24) Out and out and out the blast spreads. (COCA; Magazines; 2004)

(25) And sometimes parents just out and out doubt the teacher’s veracity. (COCA; Magazines; 2003)

(26) That is all an out and out lie and apparently his testimony has not been shaken. (COCA; Spoken; 1996)

In the figurative sense, out and out is normally used to express the speaker’s negative attitude to the action or person etc. described.

17 From the Christian hymn “The Sweet By-and-By” with lyrics by S. Fillmore Bennett and music by Joseph P. Webster)
3.10 *In and in*

The sequence *in and in* is occasionally used to describe continued concrete inward motion as in (27). This use was only found in the American material.

(27) But it’s very slow-going, as we just saw, as this storm is closing in and in. (COCA: Spoken; 2005)

As noted by the *OED*, however, *in and in* has also developed a specialized meaning in the phrases *to breed in and in* “to breed always within a limited stock” and *to marry in and in,* “to marry with near relatives, in successive generations” (OED s.v. *in and in*). An example of a concrete use of this specialized meaning is given in (28).

(28) If a farmer persists in what is called “breeding in and in,” that is, from the same stock without changing the blood, it is well known that a rapid degeneracy is the inevitable consequence. (COHA; Non-Fiction; 1853)

*In and in* is thus another example of layering where the intensified directional meaning of the sequence coexists with a metaphorical one.

4. Discussion

4.1 Semantic-pragmatic processes

The formulaic sequences in the present study are undergoing a number of semantic-pragmatic processes, which will be dealt with here under the headings metaphorization; subjectification; development of textual or discourse functions; and grammaticalization and lexicalization. Finally, iconicity as an explanatory factor will be discussed.

4.1.1 Metaphorization

Metaphorization is reflected in the tendency for the meanings of these sequences to move from expressing concrete meanings to expressing more abstract ones. This is seen most clearly in our corpus material with *on and on* and *over and over,* but is also evidenced in the
OED for by and by and through and through. On and on, which in the mid-19th century mainly expressed a physical forward motion, in Present-day English mostly refers to the abstract extension of events. Similarly, over and over has shifted from being associated with physical motion described by verbs like turn and roll to being connected with any repeated activity. On and on and over and over have thus moved from referring to movement in space to referring to continuous and repeated activity, respectively. By and by and through and through have also followed a path towards more abstract meaning. By and by originally referred to a succession of people or things, and then extended to a succession of events. Through and through referred to physical extension (like stab him through and through in (19) above) from the outset, but is now often used with abstract extension. The material thus provides typical examples of how conceptual metaphor brings about semantic change (Traugott & Dasher 2005:27–34). On and on, over and over and through and through in some instances still retain their original concrete sense in parallel with their new metaphorical senses. These three formulaic sequences are therefore, like in and in as argued in 3.10, instances of layering of meaning.

When phrases undergo metaphorization, they also tend to be affected by subjectification, as will be seen in the next section.

4.1.2 Subjectification

Subjectification (Traugott 1995, 2010; Traugott & Dasher 2005) is a tendency for meanings to increasingly become explicitly grounded in the speaker’s or writer’s perspective (Traugott & Dasher 2005:6). Traugott (2010:56–60) considers some of the ways in which subjectification has been operationalized in previous studies, concluding that it seems unlikely that there are defining criteria of subjectification that can be applied across different construction types and languages. As seen above in section 3.4, the expression of negative connotations is the clearest type of subjectification found in the present investigation. In previous studies, an increased use of first-person subjects in sentences with the item that is becoming subjectified has often been taken as a clue to subjectification, but Traugott suggests that such subjects are not necessarily linked to increasing subjectivity. Instead she argues that a shift towards negative evaluation is connected to an increase in third-person subjects. Our material provides some support for this in that there has been a slight, non-significant decrease in the use of first-person subjects with the increasingly negative and subjective on and on (dropping from
10% (14/136) in the 1900s to 6% (12/214) in the 2000s). When first-person subjects occur with go on and on it is sometimes in connection with apologies for misbehaviour, as in (29):

(29) I’m sorry to have gone on and on about these things. (COHA; Fiction; 2007)

Another criterion of subjectification mentioned by Traugott is more clearly illustrated with on and on. Traugott (2010:59–60) argues that subjectification often correlates with increasingly peripheral positions in the clause. In the COHA material there is a significant shift (p ≥ 0.05) between the 1900s and 2000s towards clause- or sentence-final position of on and on (1900s 46% (63/136); 1950s 54% (90/166); 2000s 61% (130/214)).

Subjectification has previously mainly been connected with grammaticalization, but as has been seen in the present study this type of change is also relevant in the lexicalization of formulaic sequences. This is supported by Wray’s (2002:95–100) suggestion that the use of formulaic sequences serves the goal of promoting the speaker’s interests. For example, formulaic sequences aid the speaker’s production by organizing text and can be used to influence the hearer and to manipulate information. The different sequences in this study fulfil a number of subjective functions, as for instance the general extender and on and on (see below) which is used both to structure text and to express attitude towards the items listed, and through and through which expresses a subjective attitude towards the extent of something. The main type of subjectification observed with these formulaic sequences is the association with negative subjective attitudes.

4.1.3 Development of textual or discourse functions

As already mentioned, on and on has developed a specialized text-organizing function as a general extender. This occurs when a verbless and on and on follows on, for instance, noun phrases, as in (30), or clauses or sentences, as in (31).

(30) All the great innovators are pictured there. I mean, you’ve got Dizzy Gillespie and Thelonius Monk and Mary Lou Williams and Marian McPartland and Gerry Mulligan and Gene Krupa and Count Basie and on and on and on and on and on. (COCA; Spoken; 1995)

(31) But anyhow, the Emir runs the country. A huge part of the oil production goes into his family’s pocket. He has 70 wives, and on and on and on and on and on. (COCA; Spoken; 1991)
General extenders “generalize from a preceding referent to the larger group of items to which that referent belongs” (Tagliamonte & Denis 2010:335). This is seen in (30) and (31), where (31) is typical of the and on and on instances in COCA in that it expresses negative attitude towards the items in the list, while (30) belongs to the rarer positive instances.18

General extenders (sometimes referred to as vague category identifiers (Channell 1994), coordinating tags (Biber et al. 1999) or referent-final tags (Aijmer 2002)), such as and stuff like that and or something, have received extensive coverage in recent years (e.g., Channell 1994:119–156; Biber et al. 1999:115–117; Overstreet 2000; Aijmer 2002:211–249; Tagliamonte & Denis 2010). However, it appears that and on and on has been largely overlooked. General extenders do not only serve as extenders of lists from which listeners can infer other members, but also as interpersonal expressions that express speaker attitude towards the message and the hearer (Overstreet 2000; Aijmer 2002:211–249). The general extender and on and on thus performs two of the more frequent functions that formulaic sequences typically fulfil (see, e.g., Wray 2002:52–55; Granger & Paquot 2008:41–45), since it not only organizes text, but, as seen in (30) and (31), also allows the expression of the speaker’s subjective attitude towards their utterances. As with other general extenders, and on and on can serve several interpersonal functions. For example, in the examples above, longer lists would flout the Gricean maxim of quantity that speakers should not say more than is required (as also noted by Aijmer 2002:232). Furthermore, speakers and writers negotiate common ground with their intended recipients. For instance, in (31) the hearers are expected to share the speaker’s presuppositions about other disagreeable characteristics of an autocratic ruler. In conversation, vagueness and imprecision about values and opinions relies on speakers sharing knowledge and experience. Interpersonal involvement rather than explicitness is important. In academic writing, on the other hand, writers can use a partial list and add a general extender as a hedging device (Biber et al. 1999:1045). The development of and on and on into a multi-functional general extender is thus a prime example of subjectification of meaning.

The general extender and on and on occurs in all the genres studied, as seen in Table 2.

---

18 A function similar to that of a general extender is performed by the sequence the list GO on and on, as in From Barry Bonds to Floyd Landis, the list goes on and on of athletes who are willing to sell their bodies and their souls for the sake of winning (…) (COHA; News; 2006). There were 14 instances of the phrase the list goes on and on in COHA in the 2000s. The first attestation of this sequence in COHA is from 1952.
As can be seen in Table 2, the expression of negative subjective attitudes towards the items listed is one of the main functions of *and on and on*. There is a slight tendency for *and on and on* to be more frequent in informal genres, where subjective attitudes are more often expressed, rather than in the more neutral style of academic writing, where the sequence is also the least frequent. However, there were no significant differences between the genres as regards the expression of negative subjective attitudes.

*And on and on* does not appear to be in real competition with the more established *and so on*, however, since the latter outnumbers the former by 25 to 1 in COCA. Moreover, while *and on and on* expresses negative connotations in a majority of the cases (62% (199/322)), *and so on* is more neutral with only 19 negative out of 100 random examples (this difference is significant ($p \geq 0.05$)). There is thus a division of labour between the two expressions.

### 4.1.4 Grammaticalization and lexicalization

There was only one clear instance of grammaticalization in the material, and while it is not from Standard English, it is still noteworthy. The grammaticalization of *by and by* as a marker of future time in Tok Pisin is a textbook example of a lexical item acquiring a grammatical function. *By and by* developed into *baimbai* and finally into *bai* (Sankoff & Laberge 1974:77–79; Verhaar 1995:314–315), as in *Nil kapa bai i no ros kwik ‘Copper nails will not rust easily’* (Verhaar 1995:315). The peak in frequency of *by and by* in our American material coincides with the British colonization of Papua New Guinea in the 1880s. It is therefore likely that it was a fairly frequent adverb also in BrE speech at the time, which would explain why this adverb, which is quite rare in Present-day English, could become grammaticalized in Tok Pisin.
More ADV1 and ADV1 sequences have undergone and are undergoing lexicalization.

As discussed by Brinton & Traugott (2005:37–38), it has been argued that conversion, i.e. the shift from one category to another, is a type of lexicalization. Lexicalization is probably most evident in those cases where a formulaic sequence changes word class and acquires a new idiosyncratic meaning. Among the ADV1 and ADV1 sequences, this is seen with the noun (on the) up-and-up and the adjectival out-and-out (hoodlums). Such conversion has been attested with other sequences as well, as for instance when the adverb face to face is used as a premodifier (a face-to-face meeting) and a noun (let’s have a face-to-face) (for more examples, see Lindquist & Levin 2009). Hyphenation in writing is frequent in adjectival uses and even more so in nominal ones.

However, other instances of lexicalization typically meet a number of other criteria mentioned by Brinton & Traugott (2005). Of these criteria we will discuss univerbation, non-compositionality, grammatical deficiency, lack of substitutability and simplification.

Lexicalization is frequently thought of as univerbation (Brinton & Traugott 2005:47–52). Signs of incipient univerbation can be detected in hyphenated forms, e.g., by-and-by, out-and-out used as an adjective and up-and-up as a noun. As will be seen below in the discussion of iconic variation, such lexicalized forms resist multiplication (*an out-and-out-and-out madman). The forms used as adjectives or nouns are all semantically non-compositional (ibid. p. 55), i.e. it is not possible to deduce the meaning from the parts of the sequence. These criteria are interrelated: Non-compositional sequences become stored as single units in the mental lexicon (cf. Wray 2002:128–139) and are often written with hyphens as single units. Because the meanings of the sequences are opaque, there can be no iconic variation. Non-compositionality is probably the most pervasive feature of the ADV1 and ADV1 construction, since all types except down and down have developed opaque uses. This applies both to more frequent sequences such as on and on and the rare and technical BREED/MARRY in and in, which shows that non-transparent meanings do not only develop with high-frequency items.

In the material there are two examples of emerging non-compositionality. On and on and over and over are losing their association with motion verbs, and are instead increasingly used metaphorically. While over and over co-occurs with a number of verbs denoting, for instance, communication (SAY, TELL), on and on is increasingly becoming restricted to the idiom GO on and on. Such fixing of collocates is typical for lexicalization, according to Brinton & Traugott (2005:105). GO on and on is thus also an example of formulaic sequences showing an increasing lack of substitutability (Brinton & Traugott 2005:55). There is no and
has never existed any lexical substitutability in the ADV1 and ADV1 construction itself in
that and cannot be replaced by the synonymous as well as: *she prattled on as well as on. A
factor influencing the low degree of variability in formulaic sequences, adduced by Hudson
(1998:33–36) (also discussed by Wray 2008:16), is that many such sequences are adverbs,
and adverbs are often not subject to variation either as sequences of words (by and large) or
as individual words (mostly).

It can be argued that the next criterion, grammatical deficiency (Brinton & Traugott
2005:55), is reflected in the fact that coordination of two identical adverbs by means of and
can be seen as irregular. This holds especially for sequences that are used as adjectives or
nouns. The other criteria suggested are either not applicable (e.g., lack of passivization) or not
fulfilled (lack of topicalization; on and on it went is unproblematic).

Simplification (Brinton & Traugott 2005:54) is illustrated in the ADV1 and ADV1
material by the loss of again after over and over. Expressions that are used repeatedly tend to
be reduced.¹⁹ Economy and clarity are two fundamental motivations for speakers. In the case
of the slightly tautological over and over again, it can be argued that the increasingly non-
transparent over and over can serve the same function as the longer sequence, and that
therefore the reduction of this phrase affects the clarity for hearers only minimally.

It is noteworthy that the reduction of over and over (again) was the fastest between the
1850s and the 1900s, which is the time span when the frequency of over and over increased
the most (see Figures 1 and 3 above). Since then the overall frequency of the sequence has
been decreasing and the reduction process has been slowing down. It has been shown in
numerous studies that frequency plays a role in simplification (cf., e.g., Bybee 2006), and in
this case it seems that over and over (again) went through a dynamic period of change that
coincided with its rapid increase in frequency. As the sequence was more frequently used,
speakers could simplify the form because hearers could be expected to quickly understand a
frequent sequence.

4.1.5 Iconicity

The origin of the ADV1 and ADV1 construction most likely lies in iconicity: the adverb is
repeated in order to express intensity. Such reduplications are common in the world’s

¹⁹ Reduction of lexicalizing items are mainly exemplified by phonologically reduced forms (e.g., lord < OE hlaf
‘loaf’ + weard ‘guardian’ (Brinton & Traugott 2005:47–52), but in the case of over and over again, a word is
elided.
languages and their development can be studied in pidgins and creoles. Reduplications in Tok Pisin, e.g., “convey feelings in regard to the high degree in which the modifier obtains, or to the large quantity of what is modified by the reduplicated form, or both” (Verhaar 1995:292).

Thus, for instance, bikpela bikpela means ‘in large quantities, huge, very large’.

Occasionally, the intensification is carried further and sequences of three or more identical adverbs occur in the material, as can be seen for some of the sequences in Table 3.

Table 3. Frequencies of ‘multiple’ tokens with three adverbs in COHA

<table>
<thead>
<tr>
<th>Sequence</th>
<th>1850s</th>
<th>1900s</th>
<th>1950s</th>
<th>2000s</th>
</tr>
</thead>
<tbody>
<tr>
<td>again and again and again and again</td>
<td>0% (0/207)</td>
<td>0.2% (1/424)</td>
<td>1.9% (6/322)</td>
<td>7.3% (27/369)</td>
</tr>
<tr>
<td>on and on and on</td>
<td>0% (0/19)</td>
<td>3.7% (5/136)</td>
<td>6.6% (11/166)</td>
<td>8.9% (19/214)</td>
</tr>
<tr>
<td>over and over and over</td>
<td>0% (0/138)</td>
<td>0.6% (2/317)</td>
<td>2.7% (9/330)</td>
<td>4.1% (21/518)</td>
</tr>
</tbody>
</table>

It is a reasonable assumption that the increase in multiple tokens in this written corpus is an effect of colloquialization of writing styles (Mair 2006:185–187) rather than the emergence of an entirely new way of using these sequences. Informal, more spontaneously produced language where subjective attitudes are more likely to be expressed contains more such tokens. In COCA, the highest ratios for triplicated tokens are consistently found in the Spoken sub-corpus, and the lowest in the Academic sub-corpus. Moreover, in LSAC the multiple forms on and on and on and over and over and over are significantly ($p \geq 0.05$) more frequent than in the written material from COHA, 2000s.21

Judging from COCA and the BNC, there is a slightly stronger tendency in American English than in British English to use the ‘multiple adverbs’ again and again and again and over and over and over, while the proportions are about equal for on and on and on.

Examples are given in (32) – (34).

(32) He slapped himself on the top of his head, again and again and again. (COHA; Fiction; 2000)

---

20 For all three sequences there was one relevant ‘quadruple’ instance found. The increase for again and again was significant ($p \geq 0.05$) between the 1950s and 2000s.

21 In LSAC, the frequencies of multiple tokens were as follows: again and again 33% (2/6), on and on 44% (17/39) and over and over 35% (17/49).
The hypnotic, rhythmic pounding of Roosevelt’s garbled voice droned *on and on and on*. (COHA; Fiction; 2005)

Every little thing that’s wrong is pointed out *over and over and over*. (COHA; Magazine; 2006)

More idiomatic/non-transparent cases resist triplication. While there are examples of *up and up and up* and *out and out and out* in a concrete sense, as in *go up and up and up* or *spread out and out and out*, as in (24) above, there are none in the idiomatic senses: *on the up and up* and *an out-and-out-and-out madman*.

We will now turn to some general conclusions.

5. Conclusions

In this section we will summarize the main findings of the study, answer the questions asked in the introduction, make some general comments about methodology and suggest avenues for further research.

The study has shown that in all cases except *again and again* and *by and by*, the original concrete, physical meaning is still in use together with the more recent abstract and metaphorical meanings, constituting a typical case of layering. The only case of grammaticalization found was in fact in another language (*by and by* > *bai* in Tok Pisin) but there were many sequences that to various degrees fulfil a number of the criteria for lexicalization: univerbation, non-compositionality, grammatical deficiency, lack of substitutability and simplification.

The ADV1 and ADV1 sequences behave differently as far as frequency is concerned. A majority peak before the 2000s: *by and by* and *round and round* in the 1850s, *again and again* and *through and through* in the 1900s and *out and out* and *up and up* in the 1950s. There are specific reasons for the increases of the remaining four: *around and around* and *over and over* are taking over some or most uses of *round and round* and *again and again* (but may themselves be affected by competition from other adverbials in the future). *On and on* continues its slow increase while narrowing its scope of use to the sequence *GO on and on*. *Down and down*, finally, has not developed any non-transparent meanings, is restricted to fiction and remains at quite a low frequency level.

A complex picture emerges as regards genre variation and change. The three most frequent sequences (*again and again*, *on and on*, *over and over*) are more common in fiction
than in other genres in COHA, and fiction appears to be leading the way in the increase of on and on and over and over. Furthermore, fiction is the most advanced genre when it comes to the loss of again after over and over. The simplification of this sequence is thus the most advanced in the genre with the highest token frequency. The loss of again can therefore be shown to be frequency dependent. Further studies are required to analyze why fiction writers display a preference for these sequences and why again is retained to such a high degree after over and over in speech. A more expected trend is that multiple tokens (on and on and on) are more frequent in speech than in writing. Such multiple tokens are iconic in nature, intensifying the meaning of the dual sequences, just like the dual sequences intensify the meaning of the single adverbs.

The study thus supports the argument that “ADV1 and ADV1” constitutes a construction or pattern lying behind a set of sequences which share a number of characteristics: a common grammatical form which is fixed but can be multiplied (when the sequence is used in its basic concrete sense) and in almost all cases a propensity for developing metaphorical, abstract meanings along typical paths of change. Some sequences have also developed textual functions and/or subjective meanings. Furthermore, a few have adopted new grammatical roles as nouns or adjectival pre-modifiers of nouns. It is thus reasonable to claim that these sequences make up a family of expressions, each bearing a family resemblance to the others.

Methodologically, the study has shown that for this kind of lexical investigation, very large corpora are needed. With the exception of the BNC, however, such large corpora are rarely fully comparable across time periods or regional varieties, which means that triangulation with a number of different corpora and careful evaluation of the results is always necessary.

In further research, it would be interesting to investigate the development of alternative ways of expressing the meanings expressed by the ADV1 and ADV1 sequences, and to look closer at the mechanisms behind some of the more spectacular changes (the demise of by and by; the decline of round and round). The role of genre or register in those changes also needs further investigation.

22 This generally holds true also in comparisons with spoken data from COCA and LSAC. The only exception is on and on, which is marginally more frequent in COCA Spoken than in COHA Fiction.
References


Oxford English Dictionary (OED) on line. Available at http://www.oed.com


