CLAUDIA CLARIDGE

Approaching Historical Phraseology: Patterns and Contexts

1. Introduction

Fixed and semi-fixed phrases are an important part of language: between 20% and 50% of any text may consist of multi-word units (the amount depending on the chosen definition of such expressions), according to the figures assembled from the literature by Siyanova-Chanturia and Martinez (2015, 550). While research into phraseology and idiomaticity constitutes an active branch of synchronic (applied) linguistics (e.g. Sinclair 1991; Cowie 1998; Wray 2002) and has profited from advances in corpus linguistics (e.g. Moon 1998; Biber et al. 1999), we still know relatively little about historical phraseology (cf. Knappe 2012). At least three – interconnected – perspectives are important for phraseology, namely its place in the language system, its role in actual language use, and its embedding in the mind. It is only the first two that historical studies can pursue, the sources being authentic language use (texts) and, for later periods, metalinguistic works.

In this contribution the focus will be on authentic language, which again leaves two options: using corpora or zooming in on small contexts. Using corpora may allow making statements about embedding in the language system, as frequency is at least partially linked to the institutionalization and lexicalization of phrases. Moon (1998) has demonstrated that an 18-million-word corpus can go a considerable way in shedding light on this and even on usage patterns – but also that even that size may not be enough for answering all the relevant questions. Focusing on the qualitative analysis of specific texts (as proposed by e.g. Howarth 2000) would allow making statements about the actual usage of phraseological items: their preferred contexts, their functions, their users, and their (non/semi-)creative use. This would also pay greater attention to the fact that phrases and idioms may be sensitive or even specific to certain cultural, societal, and situational embeddings (e.g. Skandera 2007). There is no guarantee, however, that the careful perusal of texts, say the plays of Shakespeare, will truly unearth all that is there. In particular, the (semantically) unobtrusive, common types such as at most, by way of, back and forth, might be overlooked, whereas a striking semantic anomaly might appear potentially idiomatic to the reader, but nevertheless be a singular occurrence based on free (and perhaps rather creative) combination of its elements. Neither can the corpus approach find everything, but it may avoid some problems connected to attention spans and introspection. It also deals with the important factor of recurrence (even rare phrases/idioms are used more than once in a speech community) by providing access to frequency information. I will therefore primarily use the corpuslinguistic method here, but in order to focus on idiomaticity combine it with suitable manual interventions and without forgetting to focus in on contexts where necessary and useful. In line with the two branches of historical phraseology proposed by Knappe (2012, 184), the modest aims of this short paper are

1 Among other problems, it is especially the prototypical idiom like kick the bucket ‘die’ which can be amazingly infrequent (e.g. only four hits in the 100-million-word BNC).
2 Such an approach would be essentially pragmatic and similar to the analysis of other pragmatic phenomena, such as Jucker’s (2000) on verbal aggression in the Canterbury Tales.
(i) to provide a synchronic snapshot of one small, probably even tiny part of Early Modern English phraseology, by concentrating on phrases based on at and by, and

(ii) to offer some insights into the diachrony of phraseology through showing similarities and differences between the historical and the modern inventory and through investigating instances of phraseological variability.

Again following Knappe, items of interest here are called "phraseological unit" (henceforth PU), which "are semantically and/or pragmatically fixed units" and "consist of two or more smaller units [...] and together do not exceed sentence length" (2012, 178-9).

This paper will proceed by first outlining previous research in this field (2) and then by presenting the data and methodology to be applied here (3). The following three sections will present selected results, namely the quantitative findings with regard to types and tokens found (4), the attested variability of PUs (5), and the distribution of PUs with regard to their functions and across genres (6). The results are discussed in the conclusion (7).

2. Previous Research

As stated above, historical research in this area is still fairly underdeveloped. Extant works are few and fall into various groups depending on their approach (cf. Knappe 2012 for a more detailed treatment).

Knappe (2004) takes a metalinguistic and historiographical view of the issue by investigating the treatment of phraseological issues in historical works on language. This gives important insights into linguistic awareness and views of historical speakers. Other works focus on phraseological data directly or combine the two perspectives. The latter is found in works which extract their data from, e.g., dictionaries, such as Bailey's or Johnson's, examples are Pinnavaia (2006) on food expressions or Claridge (2012a) on phrases based on body parts. The latter also includes corpus data (from the ZEN inter alia), and is focused on a few particular items based on content word foci; similarly Claridge (2012b) and (2013) concentrate on a small number of very specific (lexicalized) phrases (how come, as it were etc.), but with the focus on their development towards present-day English. Similar particularized approaches are found in studies dealing with (the etymology/origin of) specific idioms (e.g. Whiteword 2002), including those from particular sources such as the Bible (e.g. Crystal 2011) or in, e.g., Shakespeare (Weinstock 1966). Such studies are usually qualitative and detailed in nature. Prins (1952) focused on a large group within the phraseology of English, namely those types based on French, i.e. this work is situated within the perspective of lexical borrowing in English and covering French influence at various historical stages. The resulting dictionary was compiled through extensive (qualitative) reading, i.e. a kind of non-digital corpus. A very different corpus-approach from the above-mentioned studies is found in Culpeper and Kytö (2010), patterned on the synchronic Biber et al. (1999) approach, who focused on lexical bundles in the Corpus of English Dialogues (CED), i.e. the statistically significant co-occurrence of word-forms regardless of phraseological status, e.g. that I am, or and I will, which may be either structurally and/or semantically incomplete and do not form a lexical/idiotic unit. Finally, Howarth (2000) deals with diachronic change in the area of phraseology. He outlines three possible methods for investigating such change in history, namely: (a) targeted reading, in order to spot phrases different from the corresponding modern form...
(e.g. cast accounts vs. do accounts), (b) lexicographical data, in particular the quotations in dictionaries, and using metalinguistic expressions (e.g. proverbial) as a guide to phrases, and (c) a small-scale corpus approach, with texts on a similar topic and (apparently) manual reading. His approach is immediately focused on individual items, and relies to a large extent on researcher intuition to identify relevant data.

3. Data and Methodology

Like some of the works mentioned above, the present study is corpus-based and quantitative in its basic approach, explicitly trying to widen the data basis. But unlike some of the studies mentioned above, it tries to avoid too much a priori concentration on specific items or clearly defined phraseological subgroups. Unlike Culpeper and Kytö, the intention here is to combine the computational and idiomaticity approach by focusing on items with some kind of lexical/semantic/functional unity (cf. definition in 1). Thus the following parameters inform the study: using a large data basis and focusing on PUs with frequent function words as anchor items.

As Moon (1998) has shown, a rather large corpus is necessary to identify idiomatic patterns. Big-data approaches, however, run into natural problems in the field of historical linguistics. First, there is often not enough corpus data available, which is especially the case for the periods before 1700.3 Secondly, fully automated searching and retrieval is hardly possible, due to the high linguistic variability of earlier data. Four historical corpora are combined here to deal with the quantitative data problem, the selection comprising the Corpus of English Dialogues (CED), the Corpus of Early English Correspondence (CEEC), the Helsinki Corpus, Early Modern English Part (HC), and the Lampeter Corpus (LC).

<table>
<thead>
<tr>
<th>period</th>
<th>content</th>
<th>word count</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEEC</td>
<td>1410-1681</td>
<td>2,159,132</td>
</tr>
<tr>
<td>HC (EModE)</td>
<td>1500-1710</td>
<td>551,000</td>
</tr>
<tr>
<td>LC</td>
<td>1640-1740</td>
<td>1,193,385</td>
</tr>
<tr>
<td>CED</td>
<td>1560-1760</td>
<td>1,157,720</td>
</tr>
<tr>
<td>total</td>
<td></td>
<td>5,061,237</td>
</tr>
</tbody>
</table>

Table 1: Corpus data

Together they come to 5,061,237 words, span the time from 1410 to 1760 (thus including all of Early Modern English) and cover a variety of situations of use and genres. They represent written, informational and formal language as well as speech-related, involved and less formal language. The length of time covered may of course be taken as a problem, as there may have been considerable phraseological change. Nevertheless,

---

3 The Dictionary of Old English Corpus, for example, comprising all available Old English data, comes to 3 million words. The whole of Middle English is best covered by the Penn-Helsinki Parsed Corpus of Middle English with only 1,155,965 words. Early Modern English corpora tend to be fairly small. Only Late Modern English corpora are fairly big, e.g. the Corpus of Late Modern English Texts 3.0 with 34 million words, the Corpus of Historical American English with 400 million, and the Old Bailey Corpus with 24.4 million.
the data will mostly be treated as if it represented a synchronic snapshot, as the aim here is a diagnostic survey.

The anchor items *at* and *by*, which are frequent in themselves and known from modern English to enter into many PUs, are used as search items. Also, they do not exhibit much formal variation, which helps with the qualitative data problem mentioned above. Function words were chosen, because they are semantically and structurally versatile enough to lead to potentially very different types of combinations.

The concordance lines (produced with WordSmith 5.0) were sorted according to the first and second words on the right in order to see recurrent patterns and at the same time this focuses on non-verbal combinations. With regard to recurrence, Altenberg, for example, accepts cases "occurring more than once in identical form" (1998, 101) in modern spoken data. In order to make allowance for the historical data (the size of the corpus, the time span and the fact that emerging PUs might allow more variation), recurrence in similar form was used. Identifying recurrent patterns is complicated by spelling variants, so that not only immediately adjacent concordance lines may constitute instances of one PU. Also, as semi-fixed PUs allow inserting and transposition of items, care was taken to spot as many such variants as possible. For example, the combination of *by* with *means* (covering meanings of instrumentality and causality) turns up in different syntactic shapes in the data, among them the cases in (1).

(1) a. and others upon Hawking their Ware all about the Country, until many of them are ruined by means of their great Charge in Travelling; (LC Eca1681)
   b. for all which, beside the real Freight and Commissions staying with us for ever, we send them in Return woollen Manufactures principally, by which Means we are greater Gainers than we should be by those who took from us principally the Manufactures of India, Holland, Flanders, or Germany; (LC Eca1731)
   c. measuring Wheel, which after 'tis actually taken, must with great care be plotted down upon Paper, but not without allowing for the Variation of the Needle, and all notable Ascents and Descents with other turnings and windings, that will of necessity be met with in the way, and so by this means we shall come to know how many Miles on the Earth will answer to a Degree in the Heavens. (LC Scia1698)

These three items are widely separated in the concordance, but once a basic type of co-occurrence has been noted (here *by* and *means*, or also *at* and *ease*, cf. (10) below) a systematic search of the concordance lines for such transpositions and interruptions can be undertaken. Nevertheless, there is of course no guarantee that all variations of a given PU or a PU with too many variations have been reliably identified.

All the data was screened for phrases around *at/by* which functioned as a meaningful functional entity in line with the definition above (i.e. not just any kind of n-gram), which means that recurrent items not fulfilling this criterion were discarded (e.g. both *by all* and *at all* occur frequently, but the former is not a unit as it is only completed by a variety of nouns following *all*). Some items making a unitary impression occurred only once. Here Knappe's (2012, 185-6) listing of identification criteria was resorted to, in particular the following: a modern parallel exists, the communicational factors are present.

---

4 Nevertheless, spelling variants have been used in searching. As the focus is phraseology, not the anchor words themselves, phrases in which these are not prepositions have been accepted as well (e.g. *by the by*). In other words, it is actually the orthographical form which is the search item. Note also that the search might find PUs of which the search item might not be an integral member.

5 Thus, verbal phrases (such as *aim at, come by*) are not in focus here, as this is an area that has already been researched fairly extensively, e.g. Brinton and Akimoto (1999); Claridge (2000); Thim (2012); Rodriguez-Puente (2013). Some verbs are nevertheless part of the data, if there is a further post-prepositional PU constituent, i.e. a structure like *v + at/by + X*. 
tive function speaks for a routine formula, typical formal features of PUs are present (e.g. alliteration), or a certain degree of semantic idiomaticity/opacity is present. For checking on modern parallels, Cowie et al. (1983) and the BNC were made use of.

A few examples of items occurring only once can illustrate the selection process.

(2) a. yet he and his Son Henry II. held Charles V. for above 30 Years together at Bay, and defended their Territories against all that Charles was able to do against them; (LC, Polb1713)

b. Well, Madam forsooth, says she, Gibing at me, you would be a Gentlewoman, and pray how will you come to be a Gentlewoman? what, will you do it by your Fingers Ends? (…) Why, what can you Earm, says she, what can you get at your Work? THREE-Pence, said I, when I Spin, and 4 d. when I Work plain Work. (CED, d5defoe)

c. or that can bee willing to have this their Wisdom and Glorie to bee flatly and plainly called a Pageantrie, Mumming, or playing at bo-peep with themselves and others; and yet (it really beeing no honester) if wee cannot bear the Judgments thus one of another; (LC, Eca1652)

d. let the Devill never so cunningly hide his head, we shall know him by his Cloven foot; nor let the Asse never so craftily hide his feet, yet we may know him by his Eares: and howsoever covertly you have carried your selfe, I know you, both by head & heels: I see you are one of the New Faction. (CED, d3mwomen)

At bay has a modern parallel; it is listed in Cowie and Mackin as keep at bay and the PU occurs 310 times in the BNC, mostly indeed with keep, but in fewer instances also with hold as in (2a), and sometimes also in a non-verbal construction. But in normalized terms at bay is more frequent today. Both (2b) and (2c) might be cases of creative uses and/or allusions to PUs. In addition to (2b)'s form there are also two instances of (have) at one's fingers' ends in the data. Here, a similar PDE form exists, namely at one's fingertips (with only four instances in the BNC). The phrase in (2b) is combined with an unusual verb (do) versus more usual have, and the meaning seems tilted towards the literal, as the later answer involving spinning indicates. However, it is not inconceivable that the author here indulges in language play, activating both literalness (through by) and figurativeness by alluding to and activating the idiomatic meaning "readily/easily available," here rather "doable," by the idiom-derived NP. With regard to (2c), we find prepositionless play bo-peep both in the historical data (Shakespeare's Lear I, iv) and in the BNC (one each), so such a low-frequency PU exists. Here it is 'blended' with the prepositional verb play at, which adds semantic nuances of pretence, casualness and half-heartedness – which seems to fit well with preceding pageantrie, mumming. By head and heels (2d), finally, has an idiomatic look to it by virtue of its binomial structure and its alliteration (and cf. head over heels) and thus was considered a potential PU. However, no such form can otherwise be found in historical or modern sources. In fact, analysing the context more closely the elements of the NP are cohesively linked to the context, head with the asse's ears and heels with the devil's foot. Thus a playful, but nevertheless ad-hoc use seems to be present here, not a PU. This particular example was thus excluded.

4. PU Types and Tokens Then and Now

The search items were found in the four corpora in unsurprisingly high frequencies, namely at with 29,419 hits and by with 31,694 hits. After manual screening along the
lines explained above, 6,738 phraseological instances with *at* and 4,601 with *by* remained, i.e. 23% and 14.5% of the total occurrences respectively may be taken to enter into PUs. *At* thus seems to be the more phraseologically active of the two anchor items. *At* is also represented by more types, 152 vs. 142 for *by*, but if one considers the ratio between types and tokens *by* seems the slightly more productive one. With regard to types, *at* also leads in the present. In Cowie et al.’s dictionary (1983), there are 98 entries for *at*, and 51 for *by*, while the difference is even greater in Moon’s (1998, 90) investigation with 161 *at*-types and 55 *by*-types (third vs. seventh place in the list of preposition-headed PUs). *By* seems to be much more frequent and more productive in the past than nowadays, but this is at least partly due to the counting method: I have counted all oaths/swearing types separately (e.g. *by heavens, by St. Anne, by Christ, by my faith*) whereas treating them as one pattern would reduce the number of types by 40.

Individual types may occur very rarely or they may be repeated very frequently (few vs. many tokens), which is captured here by sorting them into frequency bands, with types occurring up to 5 times, then up to 10, 50, and 100, and finally more than 100 times in the corpora. In this regard *at*- and *by*-PUs pattern amazingly similarly, as Table 2 indicates in percent of items occurring in each frequency range.

<table>
<thead>
<tr>
<th></th>
<th>&gt; 5</th>
<th>&gt; 10</th>
<th>&gt; 50</th>
<th>&gt; 100</th>
<th>101+</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>at</em></td>
<td>44</td>
<td>17</td>
<td>22</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td><em>by</em></td>
<td>45</td>
<td>14</td>
<td>25</td>
<td>9</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 2: Token-frequency range of types (in %)

Considering all items with more than 50 occurrences *at* and *by* are highly similar (17 and 16 percent respectively), but *at* shows more of the highly frequent forms. Such very frequent forms (over 100) include both completely fixed and more flexible ones. *At all* and *at once*, for example, occur only in this invariant guise, but combinations with *command, distance, first, hands, large, last, least, length, liberty, present, rate, and time* allow various insertions and transformations. *At sb's (x) hands* may occur as *at the (x) hands of* (where x stands for various items to be inserted) and *at (the) (x) length* may be accompanied by an article and/or an adjective. Frequent invariant *by*-PUs are *by and by*, and *by way of*, while various changes may occur in combinations with *mean(s), reason, time, and virtue*. Frequent phrases with one standard open slot are *by sb's truth/troth, by God's x*.

Some individual items show a clear frequency difference in their use then and nowadays (as evidenced by the BNC). Thus, *at length* is attested at 50.6 per million historically versus only 4.5 today, *at sb's command* 43.8 vs. 0.9 and *by reason* (of) 99.8 vs. 4.1. While the decline of *at command* may very well be connected with socio-cultural change (from a more to a less hierarchical society and nowadays greater concern for egalitarianism), linguistic reasons are more likely for the other two cases. In both of these the frequency might be boosted by lack or disfavouredness of alternatives. *Because* or its precursor phrasal form *by cause* (of) was still fairly infrequent then. A synonym for *at length* in one meaning is *finally*, which was also comparatively rare, occurring only 33 times or 6.5 per million vs. 27.6 in the BNC. In contrast, a form that shows almost perfect constancy is *by day*, with a frequency of 4.7 per million then and 4.4 in the BNC. This item is hardly context-bound, stylistically un-

---

7 All types are listed in the appendix.
marked, found also in other Germanic languages (e.g. German bei Tag), and, crucially, does not have any alternatives.

At length may illustrate that the polysemy of an item may be important to look at in the context of its frequency development. The phrase then and now has two meanings, namely "finally" (as in 3) and "in detail, in a lengthy manner" (4):

(3) a. the Abbot and the Monke, [...] went and cunningly tooke him out of his graue, and carried him into a deepe dungeon, where he could see no light, and there let him lie starke naked, till such time as the potion had ended the operation, and that hee should wake. At length the Farmar awoke, [...] (CED, d1fcoble)

b. when I first propounded it to the Royal Society, 'twas look'd upon as a new thought, and somewhat extravagant, and hardly practicable, until upon hearing my explication, and the various ways how it might be reduced into practise, it was at length judged possible, and desirable to be tryed. (LC, SciA1674)

(4) a. he repeated the whole Story at length, viz. That [...] (LC, MscB1692)

b. But because I am obliged to my Good and Learned Friend Mr. Whiston, for his kind Letter and Tables upon this Occasion, I shall make use of his Leave to insert them at Length. (LC, MscB1718)

The distribution of these senses in the CED and LC versus in the BNC is as follows:

<table>
<thead>
<tr>
<th>meaning</th>
<th>1560-99</th>
<th>1600-39</th>
<th>1640-79</th>
<th>1680-1719</th>
<th>1720-60</th>
<th>BNC</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;finally&quot;</td>
<td>11</td>
<td>17</td>
<td>34</td>
<td>32</td>
<td>14</td>
<td>163</td>
</tr>
<tr>
<td>100%</td>
<td>100%</td>
<td>89.5%</td>
<td>91.9%</td>
<td>88.8%</td>
<td>77.7%</td>
<td>36.6%</td>
</tr>
<tr>
<td>&quot;in detail&quot;</td>
<td>-</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>282</td>
</tr>
</tbody>
</table>

Table 3: Sense distribution of at length

This table may bear out the assumption above, as it is the sense "finally" that dominates in the historical data, but only makes up somewhat more than a third of all cases in the BNC. This sense has decreased towards the present (we may see the beginning of the decline in the period 1720-1760), while presumably finally has increased in Late Modern English.

There are PUs in the present data which seem to have fallen out of existence at some point between 1760 and today. (5) collects a few such examples:

(5) a. The French affairs are in great decadence unlesse we can help them at a dead lift. (CEEC, marvell)

b. Syr, the soner ye be at apoynt for the occupacion and profet of the viij acres of lande it is the better, bycause it is nowe soyn with corne, and the thridde sheffe therof shulde apperteyne vnto you or your deputie (CEEC, brereto)

c. you are not to attend at Diepe in expectation of the Said Doctor longer then wednesday or thursday next at furthest. (CEEC, pepys)

d. I may saye, of this retardation by occasion of the lack of wyndes, he shall mak a shipwrack of his jornaye (CEEC, leycest)

e. And if he myght haue comme with that wynd, he wold by liklihod haue be heer IJ dayes past. (CEEC, fox)

f. Cannot I keep Christmas, eat good Chear, & be Merry, without I goe and get a Licence from the Parliament? Marry gap, come up here, for my part Ile be hanged by the neck first; must I be subject to them I never saw in my life? (CED, d3mwomen)

---

8 Five unclear or ambiguous tokens in the historical data, and ten in the BNC have been left out of consideration here.
Starting from the bottom (5f), one reason for the loss of phrases may be sociocultural change so that the extralinguistic and thus also conceptual area necessary for the (literal) background of the phrase is no longer available (Burger and Linke 1984); the kind of capital punishment referred to is simply not practiced any longer. Burger and Linke (1984) list further possible reasons for loss. Among them is the competition of semantically similar phrases which may account for (5c, d, e). By occasion of has been superseded by because of/ by reason of, by likelihood by in all likelihood and at furthest by at most. Another reason may be the loss of a semantic component of one of the elements, e.g. in at a point 'concluded, agreed' (5b) point has the sense 'resolution, decision' which nowadays is obsolete. This same reason as well as sociocultural change may work together in the first phrase here (5a), as dead lift is "the pull of a horse etc. using his utmost strength at a dead weight," and figuratively the "position in which one can do no more, an extremity" (OED, latest instance 1812, s.v. dead lift). Of course none of these reasons is compelling, as a number of phrases have survived despite of them (witness the ubiquitous idiom example kick the bucket).

Instead of only looking at what is found in the historical data, one could also ask what is not there. Some well-known and common modern phrases are not included in the data, e.g. at long last, at short notice, at the double, at face value, and by and large. All of them occur in the BNC but in very low frequencies (from 0.4 to 4.8 per million), which might indicate that the size of the present data basis is still too small for this investigation. This may indeed be true for at long last and by and large, which according to the OED were in existence in the period (1523 and 1707 first attestations, respectively). For the remaining three, however, the period of investigation may simply be too early, if the OED attestations of 1785, 1865, and 1883 should indeed be the first ones and cannot be antedated.

5. PUs: Variability and Change

Many of the phrases investigated show internal variation, which may of course be an indication of incipient phraseological status and of change. Variability may concern the form of one of the central constituents, as visible in (6). The modern form is be at sb's disposal, with an open slot for a possessive. This form, as in (6a), is more frequent in the EModE data than the forms illustrated in (6b-d), but only barely so with 5 instances versus 4 each for dispose and disposing as well as 3 for disposition.

(6) a. A Tradesman's Daughter, her small Portion about 180l. and all in her own Hands, or at least in her own power, her Parents and Friends all Dead, and therefore at her own disposal, (LC MscB1692)

b. he ys sensles that conceaveth not that yf the king of Spain had these countreys at his comandment, lett hir majesty have the best peace that ever was or can be made, and wee shall find, as the world now standeth, that he wyll force the queen of England and Englond to be at his dysposytion, (CEEC, leycest, 1586)

c. I say (my Lord) that seing my father grants, I will not gainsay, what his age thinks meet, I do appoint my selfe (my Lord) at your dispose, (CED, d1cknave, 1594)

d. both of them beeing pleased to express there favor so farre as to give us a for ievells, and in monie, all which with our selves we shall bee readie to cast downe at your La95feete, and bee holie at your La95disposing, (CEEC, Cornwal, 1631)
The form *disposal* is only found from the mid-17th century onwards, mostly in LC and once in CEEC, but there is still competition, in particularly with *disposing*, in the 17th century. Thus the consolidation of a phraseological form can be seen in the data. Most historical forms contain a possessive pronoun (like 6a-c), which is also the most common modern realisation. (6a) shows a further slot to be filled (*own*), which is also still sometimes found in the modern phrase (e.g. BNC *at its own disposal*, CMT 205). *Being at disposal* then (in any of its forms) as well as now can refer to people (6c) or things (6a), with the former more common in the past.

Another variable case is found with *at least*, which occurs with or without definite article and with three nominals, *least, leastwise* (one or two words) and *least way(s)*. The standard modern form clearly dominates in this case, but competition is found until the early 18th century.

<table>
<thead>
<tr>
<th></th>
<th>CED</th>
<th>CEEC</th>
<th>HC</th>
<th>LC</th>
</tr>
</thead>
<tbody>
<tr>
<td>least</td>
<td>121</td>
<td>343</td>
<td>73</td>
<td>285</td>
</tr>
<tr>
<td>leastwise</td>
<td>6</td>
<td>8</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>least way(s)</td>
<td>-</td>
<td>5</td>
<td>6</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 4: Forms occurring in second place in *at LEAST*

Both complex forms are found in modern dictionaries, but marked as 'informal' by the Longman DCE, for example. *Leastways*, but not *leastwise*, occurs in the BNC (13 hits), notably without *at*, mostly in fiction and in the company of colloquial and/or nonstandard features, cf. (7) for example.

(7) She went very red, and seeing he'd scored, Will added slyly: 'I ain't going to make trouble, leastways not if th'bist sensible. (BNC C85 1312)

There is no informality issue visible in the older data, as the occurrences (8a-c) in an educational text, a trial and a letter by Thomas More show. While the majority form from the past thus made it into the standard, the other forms are apparently marked minor options today.

(8) a. And hit shal be no reproche to a noble man to instruct his owne children, or at the least wayes to examine them, (HC, ceced1a)
   b. My Lord, this must tend to the preventing all manner of Justice; it is against all common Sense or Reason; and it never was offered at any Lawyer before, as I believe, at leastwise never so openly, (CED, d4trookw, 1696)
   c. and at the leaste wyse I remember well, that of those poyntes which you call now newly to your remembranace there was none at that tyme forgotten. (CEEC, More, 1543)
   d. and all their Designs to destroy our Religion, have either been begun, or, at least, carried on and promoted by that City. (LC, RelA1730)

Regarding the article (as in 8a and c), the BNC returns 25,058 hits for the form *at least* and 88⁷ for the sequence *at the least* (c. 0.3% with article). There is a slight meaning difference between the two (the article reinforcing the literal meaning of *least*), which, however, is not present in all historical examples, cf. (9), where *at least* would be more appropriate in modern English:

(9) Let it suffice thee, (Mistris Page) at the least if the Loue of Souldier can suffice, that I loue thee: (HC, ceplay2a)

---

⁷ In fact, there are 113 in total, but 25 are part of larger, non-idiomatic NPs.
The article is thus semantically often unmotivated, but nevertheless present to a considerable extent (from 3% in LC via 12% in CED to about a third of all cases in CEEC and HC). As the article is the default with superlative forms its presence in a considerable number of cases shows the phraseologism to be still more at the open, regular end of the spectrum.

The modern phrase *at sb's beck and call* is a low frequency form, with only 0.4 hits per million words in the BNC. In the historical data it does not appear at all, which might at first sight be attributed to too small a database. It does, however, occur in two different shapes, namely with one of the two nouns only, as shown in (10).

(10) a. No, no, you are out, my Lord, said Gerardo for she's a Bawd of better Quality, than to be *at the beck of* a Footman. (CED, d4ftrage)
   b. and me thought I was something impatient that the Lord should surprize us, getting us into a ship at his call for his service, and then to drowne us, as if we were Parricides or hainous Malefactors, (LC Relb1650)

It is found with *beck* three times, once each in CED, CEEC and LC, and four times with *call* or *call*ing), one each in CED, CEEC and two in LC. Returning with this knowledge to the BNC reveals that these single forms are still possible today, with three attestations each (versus 42 for the longer form). Nevertheless, at some stage, probably after the time of the corpora used here, the two phrases must have merged to form the larger one. The OED has a first quote with the full phrase dated 1869 (s.v. *beck* n.2), which seems a bit late given the age of the smaller forms. The merger presumably has to do with necessary reinforcement of *beck*, as its independent meaning was bleaching. As binomial frames are well established in the language, the form of the reinforcement seems highly natural. A similar case may be *by fair means or foul*; in this case, however, *by fair means* is attested in the corpora investigated (3 CED, 2 CEEC), but not *by foul means*.

Another form of variability concerns not the main building blocks of a PU but further modification possibilities. The form *at ease*, for example, appears to allow only possessive pronouns today (as visible in the BNC). Fourteen percent of the historical items, in contrast, do contain some other modification, as (11) shows. Thus, the phrase can again be seen as less fixed in the past.

(11) a. I shall be *at some harte's ease*, which at this time I have neede of ... (CEEC, cornwal)
   b. That the Rider is *at extreme ease*, and may, for ought I know bee able to travel in it 3 days together, night and day. (CEEC, petty)
   c. and it is certain yt if ye Revenue bee made good his Mat' will bee *at great ease* for all concernments here. (CEEC, essex)
   d. then was I *at Much Ease* (CEEC, pepys)

The reasons for the variability may or may not be phraseological in nature. In the case of *at sb's disposal* the anchor item is also highly variable outside this frame, the speech community not having settled yet on a clear form-function pairing. Once this had happened (sometime past 1700), the fixation apparently also applied in the PU. With *least/leastwise/leas*t ways, in contrast, the competition of these forms in the past only occurs *within* the PU. *Least*, being a very frequent non-PU form, however, will have supported the phraseologization in the shape *at least*. In the last example in this section, *at ease*, the general modification patterns seem to play a role as well, as *heart, great and much* often modify *ease* outside the PU. As long as the same general collocation patterns apply one may even be tempted to see *at ease* as only border
phraseological. Thus, assessing the status as of a PU one will always have to pay attention to the behaviour of its component items in the language at large. Whether variability of the anchor item (least) or the presence of additional items (ease) has a greater slowing down effect on phraseologization will need to be investigated further on the basis of data of a longer time period.

6. PUs in Context

The distribution and normalized occurrences in the different corpora investigated is also of interest here, as shown in Table 5.

<table>
<thead>
<tr>
<th></th>
<th>CED</th>
<th>CEEC</th>
<th>HC</th>
<th>LC</th>
</tr>
</thead>
<tbody>
<tr>
<td>at</td>
<td>14.1</td>
<td>10.3</td>
<td>10.7</td>
<td>15.6</td>
</tr>
<tr>
<td>by</td>
<td>9.3</td>
<td>7.7</td>
<td>12.0</td>
<td>7.3</td>
</tr>
</tbody>
</table>

Table 5 PUs with at and by in the four corpora (per 10,000 words)

While not evenly distributed, the PUs are well represented in all the corpora, indicating that the chosen anchor words provide a wide spread of types, which are of use in many different contexts. Interestingly there is no clear pattern sorting apart speech-related from written or public from private language, although one might expect that clearly speech-related language shows more PUs. By ranks highest in the HC and at in the LC, both of which represent the more written end of the spectrum. The majority of the PUs are of the ideational type (cf. below), to a lesser extent of the interpersonal type, which may correlate with greater use in writing.

We can zoom further into these corpora by looking at specific genres. For illustration I have chosen law and drama, which allows comparison within one domain but across speech and writing (treatises vs. trials) and within the speech-related type but differences in topics and authenticity (trials vs. plays).10 Given the different text sizes we find the fewest types in HC-law (19), followed by HC-trials (32) and CED-comedy (105). There is incomplete overlap, as only five types are shared by all three categories (at sea, at the least, by means of, by reason of, by way of). With the exception of at sea, all of these show little propositional restriction, doing more structural and interpersonal work instead. A further 23 types are common to two text groups:

- law & trials: at sb's pleasure, by experience;
- trials & comedy: at all, at first, at first sight, at length, at once, at sh's hands, at peace, by and by, by chance, by force, by God, by no means, by name (of), by sh's leave;
- law & comedy: at all times, at liberty, at most, at sh's request, by force of, by stealth, by virtue of).

Again, none of these are very specialised in nature, e.g. doing something by stealth can be used for describing different situations, whether it is love affairs or smuggling. One may perhaps find more genre- or situation-specific types among those only found in one of the three genres. Interestingly, none of the remaining five types in legal treatises fits the bill (at large, by occasion of, by the grace of, by the space of, by water). Three of the eleven items found in trials may be more relevant in this respect: (12a) arises immediately out of the (originally purely) physical context in

---

10 HC legal treatises: c. 38,000 words, HC trials: c. 48,000 words, and CED comedies (periods 1-4): c. 190,000 words.
court, (12b) acquires more specialised meaning in this context, and the elaborate speech act in (12c) only makes sense in a legal system that permits royal pardon.

(12) a. My Maisters of the Jurie, you haue to inquire, whether Sir Nicholas Throckmorton Knight, here Prisoner at the Barre be guiltie of these Treasons, (HC cetr1)
   b. There is a Law of two sorts of Accusers, one of his own Knowledge, another by Hear-say, (HC cetr2b)
   c. It is true we did consult at Drewry house, about securinge the E. of Essex his access free from impeacht and that for noe other end, but to pstrate o' selves at her mat' ffeete, humble submitinge o' selves to her m'cy, (HC cetr2a)

The CED comedies contain 79 PUs not found in trials and treatises, the high number of course also due to the larger textbase. What is noteworthy is that eighteen of those contain a slot for personal reference (e.g. at sb's word), of which eight explicitly have interpersonal/politeness implications (e.g. at sb's service or call). 16 realise swearing formulae, e.g. by the masse, by this light, by gogs. These highlight the uses of PUs in interactive and involved contexts. As drama is action unfolding in time it may not be suprising that there are also a number of time-related PUs (e.g. at the cock's crow, at last, by that time) and PUs referring to states and actions (e.g. at dinner, out at heels, come at unaware, go together by the ears). Comedy is also the category where one finds textual uses as in (13).

(13) a. I hope you will say so, when you have heard all George; but by the way your late stock being spent, here are ten peeces towards a supply. (CED, d3cbrome)
   b. I assure your Ladiship, though my wife can boast as great and noble friends I thank fortune, as the wife of any Tradesman that carries a head in the City, (but that's by the by) yet I came of a better house, and am a Gentleman borne, (CED, d3cbrome)

The previous points highlighted an important aspect about the contextual use of PUs, namely the function they fulfil. Using the common three-fold Hallidayan distinction ideational/interpersonal/textual (also used for idioms by Fernando 1996), one can sort all historical items accordingly. First, ideational phrases are those that have pronounced conceptual content and are used to contribute to propositions about the state of affairs in the world, for example at low ebb, at first sight, at hand, at the helm, at once, fall by the ears, by fits, by heart, by turns, etc. Secondly, interpersonal phrases relate more closely to the speakers, their state of mind and emotions as well as their interactions with the addressee, e.g. at a certainty, at a guess, by your favour, by my troth. Textual PUs, finally, function in making explicit propositional and textual connections and structures, e.g. at first, at the same time, by the way, by reason of. Classification has been done here on the basis of the meanings and prototypical functions of the types, without checking the applicability of the classification for every single token. A small margin of error is therefore possible.

A classification of the at and by types reveals a rather different functional profile (Figure 1). At-PUs function much more prominently on the ideational level, i.e. they are used predominantly to package content adverbials (e.g. indicating manner, location, time etc.). This is to a lesser extent true for by, which may have to do with the fact that by is the less versatile preposition with a more restricted range of meanings. Textual PUs are rare with both anchor words, but this is unsurprising. A restricted, but repeated number of versatile phrases can do a lot of textual work (a token approach would show a higher percentage here). The most noticeable difference between the
two is found with regard to the interpersonal function, where *by* is much more active. This deserves a closer look.

![Figure 1: Phraseological types and their functional distribution](image1)

With regard to the present data, the interpersonal group can be further subdivided into politeness PUs (e.g. *by your leave, at your service*), stance and evaluative PUs (*at best, at least, by far, by all likelihood*), and swearing (e.g. *by the fear of God, by Bacchus*).

![Figure 2: Interpersonal functions subclassified](image2)

The full three-fold distinction is only necessary for *by*, however. *At* here is most active in expressing speaker stance and to a lesser extent in marking politeness. *By* shows little politeness activity, somewhat more evaluative uses, but specialises in particular in swearing and expletive phrases. There is no comparable modern data, but one can assume that the distribution would look rather different. Swearing in this particular form has mostly disappeared, for example, and politeness is expressed more indirectly, less in the deferential form typically realised by such PUs as *by your leave, at your command* in this data. This illustrates Burger and Linke’s (1984) point of some PU’s communicative benefit being lost through sociocultural change. The modern distribution may be shifted towards more ideational uses in general and toward more evaluative uses in the interpersonal realm.

### 7. Conclusion and Outlook

The corpus approach pursued here cannot be labelled ‘big-data,’ neither with regard to the corpus size nor to retrieval methodology (namely non-automatized). However, it may be called ‘big’ in other senses: Through very frequent and versatile anchor words as search items it produced as many as 61,113 data points for further investigation. These were reduced to 11,339 relevant hits by careful manual sifting of the data, pay-
ing attention to the possibility of long and short forms, variable items, and morpho-syntactic transformations. An automatized approach would have missed many of the types listed in the appendix even in the case of a lemmatized corpus. Thus, this kind of investigation has made it possible to (i) provide a fairly comprehensive inventory of a small section of EModE phraseology, (ii) compare this to the present-day situation in some respects, (iii) present a functional profile of this whole PU group, and (iv) focus in on more specific textual environments by using the genre categories of corpora as an entry point. All of these steps actually required delving into the (micro-)contexts of the PUs (albeit without reading whole texts in detail), and the availability of the kind of background knowledge one only acquires based on extensive reading of EModE texts. Nevertheless, the next step after (iv) and based on all these steps, will have to be the detailed functional analysis of PUs in their situational and textual contexts.

Works Cited
1. Corpora
A Corpus of English Dialogues 1560-1760. Compiled under the supervision of Merja Kytö (Uppsala University) and Jonathan Culpeper (Lancaster University), 2006.

2. Secondary Literature
—. "The evolution of three pragmatic markers: As it were, so to speak/say and if you like." Journal of Historical Pragmatics 14.2 (2013): 161-184.


**Appendix**

**PUs with at:**

- at adventure: at (x) ease
- at (a) (sb's) advantage: at (sb's) ease
- at (adj) age: at a (low) eaves
- at all: at sb's elbow
- at (the) anchor: at an (sb's) end
- at arms: at eye
- at sb's back: at an exigent
- at (det) bar: at (sb's) expense(s)
- at bay: at sb's faith
- at the (sb's) beck of: at (the) farthest/furthest
- at the beginning: at sb's feet
- at (the) best: at the foot of sth
- at a/one blow(s)/stroke: at one's fingers' ends, have
- at (the) bottom: at (the) (very) first
- at bo-peep, play: at (the /sb's) first sight
- at (the) break of day: at fisticuffs
- at sb's/a call(ing): at (the) full
- burn the candle at both ends: at a guess
- at a (x) certainty: at gaze
- at chance: at (some) hand
- at a clap: at sb's (superior) hands
- at any cost: at sb's right hand
- at (sb's) command(ment): at (the) first (best, dearest)/second hand
- at (sb's) cost: at hazard, run~
- at the cocks crow: at (the, sb's) heart
- at first dash: at heels
- at this/a day: at sb's heels
- at the dead of night: at the height of sth
- at a dead(j)lift: at the helm
- at death's door: at home
- at default: at a horse tail, hanged
- at demand: at (adj) interest
- at sb's (special) desire: at issue
- at difference: at (the) large
- at dinner: at (the) last
- at (sb's) discretion: at sb's last gask, be
- at sb's dispose/disposal/disposing/disposition: at law
- at (det.) (adj) distance: at (the) least (way)
- at sb's (very) door(s), lie/lay~: at (the) leastwise
at a/sb's (adj) leisure
at (the) (adj) length
at (full) liberty
at sb's liberty
at (the) long run/renning
at (a) (adj) loss
at meat
at sb's mercy
at (the) most
at naught/nought, set
at (a) (sb's) need
at (the) odds
at once
at one
at (the) pains (of)
at par
at a (adj) passe
at peace
at sb's (adj) peril
at a period
at a pinch
at play
at (sb's) pleasure
at plough
at a (adj) point
at the (very) point of
at (det) present
at any price
at quiet
at rack (and manger)
at random
at a/det (adj) rate
at sb's request
at rest
at the root
at school
at sea

PUs with by:
by (an) (adj) accident/ce
by art
by authority
by Bacchus
by the bowels and blood of Jesus
by(-)ther(-)by
(up)on/of/in the by(e)
by(-)and(-)by
by birth
by the blessing of God
by boat
by(-)cause
by (adj) chance
cheek by jowl
by choice
by Christ

by coach
by cock (and pie)
by (the) colour(s) of
by consequence
by day
(by) open deed
by (sb's) default
by degrees
by (the) dint of
by (sb's) discretion
by the sb's ears, fall/go/set/wolf-
by earth
by the end
by error
by (sb's) example
by experience
by my (sb's) fait(h)/fay by (the) mean(s) (of)
by far by all (the) mean(s)
by my father's skin by any mean(s)
by sb's favour by no mean(s)
by the fear of God by miracle
by my fidelity by (a) (adj) mistake(s)
by sb's finger's ends by (the) mouth
by force by much
by fair means (or foul) by the naked eye
by fits by (a/sb's) name (of)
by(-)gar by nature
by (all) the gods by the/shb's neck, hanged-
by god / gosse (almighty) by night
by god's / gogs X by (the) occasion(s) of
by the grace (of god/Lord/Jesus) by sb's patience
by great by piece-meal
by god and the country by (the) (num) post
by god and my peers by (a) proxy
by guess by (the) reason (of)
by (the) half by right
by halves by rote
by all Hallows by S. Anne
by this (good) hand by S. Davids mouth
by the (one) hand by S. George
by strong hand by S. Mary
by (adj) hap by sea
by head and heels by score
by the head and shoulders by (sb's) side
by hearsay by (the) sight
by heart by ship
by heaven(s) by (sb's) soul
by the heels, lie/lay by the space of
by hell by stealth
by (the) help of by steps
by my honesty by the sweat of her brows
by my honour by sb's/the sword, fall
by hook and/or by crook by sword and fire
by horse by det time
by James by time(s)
by Jupiter by trade
by Jove by then
by ('r/our) Lady by the throat
by the Lady Mary by sb's truth/trot(h)
by land by turns
by law by the virtue of
by sb's leave by Vulcan
by my life by my vusse
by this (good) light by (the) want(ing) of
by this/shb's light by water
by (all) likelihood /~ness by the way
by little and little by (the) way(s) of
by long by word (or deed)
by the/shb's Lord by word of mouth
by the long run by the world
by the (adj) masse by yea and (by) no/nay