Multi-word Units of Meaning in 16th-century Legal Scots

Joanna Kopaczyk
Adam Mickiewicz University, Poznań, Poland

1. Defining multi-word units

A lexeme, as an entry in the mental lexicon of the user, constitutes a unit of meaning together with all its potential word-forms. It may be morphologically simple or morphologically complex, made of two or more elements, which is a definition valid for diachronic studies as well (Nevalainen 1999: 335). In order to study the semantics of such complex multi-word units (henceforth MWUs), one should first agree to treat them as lexemes, even if built of more than one lexical morpheme.

A lexical unit is the outcome of a lexicalisation process. The question arises whether MWUs are effects of word-formation or, indeed, phrase-formation. If the latter is true, can we then refer to the result of phrase-formation as lexemes? In his seminal book on word-formation, Bauer (1983) does not include MWUs other than compounds. There is a section, however, on the so-called phrase compounds (7.2.1.9). These consist of a phrase which “seems to be involved in the formation of a new word”, thus bringing lexicalisations of syntactic structures into the area of word-formation.

Bauer (1983: 42–61) sees lexicalisation as the final stage of a process affecting multi-morphemic creations and fixing them in the lexicon, “[a] diachronic shift from nonce formation to apparently arbitrary and unmotivated lexeme”. Thus a string of words and morphemes becomes fixed as a part of language and develops its own specialist meaning or function. Baldwin (2006) links lexicalisation to non-compositionality, saying that “non-compositional MWEs [multi-word expressions] will be lexically listed”, having a separate entry in the mental lexicon of the user. Bauer (1983: 49) also notes that semantic opacity is “not a necessary prerequisite for lexicalisation”, even though many lexicalised forms are opaque (see the discussion of opacity in legal language in section 2. below).

Drawing on Bauer, Moon (1998: 7–8) proposes the following criteria for whether a string of words can be classified as a fixed unit: (1) institutionalisation, whereby a string becomes recognised and accepted as a lexical item in a language, which is in principle a diachronic process; (2) lexicogrammatical fixedness, as there are often restrictions on the syntactic employment of such a string; (3) and finally, non-compositionality, which Moon sees as a semantic or pragmatic criterion. In the case of pragmatic non-compositionality, the string would be decodable compositionally (semantically) but at the same time it would serve a special discourse function. This is an important point in the context of the present paper, which will be referred to in the discussion of the data.

One of the categories included in the discussion of fixed expressions in Moon’s monograph is formulae (1998: 21–22): strings which are specialised pragmatically, while generally compositional semantically. Binomials, the MWUs chosen as the focus of the present paper, will be included under simple formulae, which are often iterative or emphatic. However, as pointed out by Moon, there are problems with non-compositionality. Essentially, a decision whether the meaning in a particular string is or is not transparent and divisible is synchronic and idiolectal in nature.

Not all scholars are ready to accept that a multi-word unit can be fixed without being non-compositional. In an influential definition of idioms, proposed by Fernando and Flavell (1981: 17), the meaning of an idiom is not the result of the compositional function of its constituents. This criterion remains the most important aspect of idiomaticity, stressing non-compositionality as a key-feature of

1 I would like to thank the organisers of the Hel-Lex2 conference for their support, as well as the anonymous reviewer for helpful observations and suggestions. All the remaining deficiencies are my own.

2 Hudson (1999: 8-9) gives a special emphasis to syntactic and collocational restrictions in her definition of fixedness.

3 Interestingly enough, in artificial languages such as Esperanto, non-compositionality does not occur (or rarely, if it does) (Dasgupta 1993).
idioms. However, according to Grant and Bauer (2004: 45), non-compositionality in an idiom must be combined with institutionalisation and, ideally, fixedness. The issue of fixedness does not seem to be so straightforward. In her corpus study of fixedness, Moon (1998: 120) notes that around 40% of fixed expressions have some lexical variation, while 14% have more than two variants. Therefore, even though fixedness is a potentially important feature of MWUs, it cannot be a decisive factor in what to treat as a MWU.

As intimated above, MWUs encompass a variety of constructions. In the present paper, the focus is on legal language and the semantic nature of MWUs typically found there. The most characteristic and intriguing type of such constructions is a binomial pair. Bhatia (1994: 143) defines a binomial as two coordinated items, joined by a syntactic link, between which there is some semantic relationship. The nature of this relationship will be explored below in section 3.

2. Legal language and the nature of its lexicon

The relationship between the function and meaning of MWUs, especially binomials, in legal language still requires discussion and further consideration. The first question is, why multi-word units were becoming part and parcel of legal registers. The next thing is how to analyse the semantic structure of these units. The final issue considered here is how the lexicalisation of multi-word units interacts with the requirement of transparency of meaning in legal contexts.

Legal language is known for its neutrality and generality (Rissanen 2000: 120). Such texts strive to avoid subjective attitudes and regional marking. In their transmission care is taken not to alter either the language or the message, which leads to the conservative character of legal discourse and the rise of formulaic expressions (Meurman-Solin 2004: 31). The expressions used in legal texts become formulaic and fixed, and may be transmitted to other text types and styles on the wave of standardisation. Alternately, during the periods of standardisation, laws drew on structures found in other genres, “decontextualised and deregionalised them” (Rissanen 2000: 121) and in this way contributed to the formation of standard language. Rissanen provides an exhaustive account of the role and behaviour of English legal writings, statutes and other documents in the development of Standard English. One can assume that legal Scots played a similar role, although this relationship between standardising Scots and legal language has not been systematically addressed yet.

Genres and registers do not function in a void. Since they interrelate and draw upon one another, it is advisable to look for general trends and tendencies in other contemporary writings in order to get a more extensive picture. In the sixteenth century, there was a certain predilection for structural parallelism and recursive constructions, not only in special-purpose genres but also in literature. Adamson (1999: 568) mentions apposition as a structural principle for literary texts noting that it was explicitly promoted by the advocates of good style. The author of Brevissima Instititio (1567), William Lily, treats apposition both as a rhetorical and syntactic category. He defines it as a “direct or indirect conjunction of two substantives in the same case, one of which is explained by the other” (translated from Latin, Michael 1970: 136), which in some way reflects the practice of producing binomial strings, such as those found in legal language. Adamson also draws attention to the principle of perspicuity, advocated by the English Renaissance theorists (1999: 600). This principle expects mutual intelligibility of speakers and that “language should act as a transparent window on the world”, shunning referential opacity, much as is required of legal language.

Authors of legal texts seem to recognise the need to keep the language as unambiguous as possible. Rissanen (1999: 192) pointed out that this struggle for clarity results in verbosity and an excessive use of expressions close in meaning, covering all possible interpretations. This is a paradox, a clash of interests present in legal genres. The striving to clarify and hone meaning with greater

4 On closer inspection, early legal texts show not only coordinated pairs but also triplets, quadruplets or longer fixed strings. In my larger post-doctoral project, I will be extracting these from corpora and applying the analytic methods tested on binomials in pilot publications, as well as along the lines of the present paper.

5 My interest in binomials was fuelled by the fact that such expressions are indeed mentioned as characteristic of legal language but their employment had not been systematically studied in corpus material. In my earliest research on binomials (Bugaj 2006), I was looking at etymology as a potential driving force behind the composition of binomials, along the lines of the “interpretation hypothesis” (Mellinkoff 1963). It turned out that etymology does not seem to play a role in the selection and alignment of binomial constituents.
precision may be the reason for expanding single lexical items to MWUs, such as bi- or tri- or multi-nominals.

MWUs and lexical fixedness in general are frequent in legal discourse, which does not go unnoticed by the authors of all monographs and studies devoted to this theme (Mellinkoff 1963, Koskenniemi 1968, Hiltunen 1990, Goodrich 1987; Bhatia 1993, Gibbons 1994, Tiersma 1999). Binomials in particular, such as to grant and to give, acts and statutes, counsale and help, constitute one of the text-type markers. Why then should they be of high frequency in legal texts? Are they an answer to the clarity-verbosity paradox, or is it just a “worthless doubling of synonyms”, to use a quotation from Mellinkoff’s seminal study on legal language (1963: 349)? Koskenniemi says in her early study (1968: 78): “In legal language a double expression is generally employed for the sake of precision and not merely for rhetorical emphasis”. She suggests that, if there are two coordinated words, they convey “a real alternative” in this type of discourse, unlike in literature. Binomials in literary texts would, in turn, serve as a stylistic figure, overlapping with various rhetorical devices (see above), and they also imply emphasis (1968: 112). It remains the purpose of the present paper to investigate the role of meaning in the use of binomial expressions in legal texts.

3. The semantics of binomials in Scots legal texts

The puzzling feature of MWUs, especially in legal texts, is the clash between their fixed form (the intended non-compositionality) and their function, which presupposes the importance of the individual meanings of the constituents (the intended compositionality). As mentioned above, the most characteristic type of a multi-word expression in legal discourse is a binomial pair. Looking at binomials, there are two aspects of meaning to consider: the meaning of the unit and the meaning relations between the constituents. As a point of departure, one could choose Malkiel’s influential paper (Malkiel 1959: 138), in which he points out two possibilities: the meaning of a pair may be a compositional sum of A and B, or an extended non-compositional new meaning, often symbolic or metaphorical. The first idea runs counter to the definition of a multi-word unit just mentioned, in which the whole expression should be lexicalised and fixed in form and its meaning non-compositional. Semantic opacity should not, however, be the case in legal discourse because legal language must not be vague for obvious pragmatic reasons (see Rissanen’s paradox of verbosity for the sake of clarity). It follows then that in LSP contexts opacity should not be the underlying cause of phrasal fixedness.

The data to illustrate this paradox comes from samples of legal material, available through The Helsinki Corpus of Older Scots (Meurman-Solin (comp.) 1993, henceforth HCOS). The texts (burgh and parliamentary records, statutes, see Table 1) fall into the legislative category, according to Bhatia’s three-fold categorisation of legal writing (1983).

6 Scots as a standardising medium of official documents in Scotland provides material just as interesting as texts from south of the border. In earlier studies, I drew attention to the fact that in the fifteenth and sixteenth centuries both languages were at a similar stage of standardisation (Bugaj 2004: 20, 28-30; 2006). Apart from this unifying factor, however, English and Scots were on a path towards a standard in two different countries, two different cultures, and drawing on different resources. Hence, both languages deserve attention individually, not simply in a comparative perspective.

7 Consider Hudson’s remark on the lack of earlier research in this area (1998: 32).
Table 1. Corpus characteristics (HCOS)\(^8\)

<table>
<thead>
<tr>
<th>Texts</th>
<th>Dates</th>
<th>Word count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acts of Scottish Parliament</td>
<td>1525–1555</td>
<td>30844</td>
</tr>
<tr>
<td>Stirling burgh records</td>
<td>1519–1529</td>
<td>11019</td>
</tr>
<tr>
<td>Aberdeen burgh records</td>
<td>1519–1556</td>
<td>10557</td>
</tr>
<tr>
<td>Carnwath Barony book</td>
<td>1523–1524</td>
<td>1748</td>
</tr>
<tr>
<td>Edinburgh burgh records</td>
<td>1540–1551</td>
<td>3140</td>
</tr>
<tr>
<td><strong>Total word count</strong></td>
<td></td>
<td><strong>57308</strong></td>
</tr>
</tbody>
</table>

Studying binomials in legislative texts, I have realised that the relationship between constituents A and B may rest on some semantic difference, e.g., complementation or antonymy, or may indeed be an instance of semantic repetition (Bugaj 2007). Semantic repetition, distinguished by Wang (2005: 510) from lexical and syntactic repetition, results from “a concept [being] repeated in one form or another”; in other words, it is a result of reiteration or relexicalisation of A by means of B, or the other way around. Consider the examples in Table 2.

Table 2. Examples of binomials (HCOS)

<table>
<thead>
<tr>
<th>NOUNS</th>
<th>VERBS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>assent and consent</strong></td>
<td><strong>anssurit and denyt</strong></td>
</tr>
<tr>
<td>bigging and completing</td>
<td><strong>clenge and make clean</strong></td>
</tr>
<tr>
<td><strong>captain and skipper</strong></td>
<td>big and reparse</td>
</tr>
<tr>
<td>dekin and craftsmen</td>
<td>chosyn and sworne</td>
</tr>
<tr>
<td>fredome and occupacioun</td>
<td>cum and gang</td>
</tr>
<tr>
<td>god and halikirk</td>
<td>dowblit and triblit</td>
</tr>
<tr>
<td>hen and poultrie</td>
<td><strong>fit and remove</strong></td>
</tr>
<tr>
<td>law and practick</td>
<td><strong>insert and registrat</strong></td>
</tr>
<tr>
<td><strong>prise and distrenye</strong></td>
<td><strong>purportit and bure</strong></td>
</tr>
<tr>
<td>peax and weir</td>
<td>takin and payit</td>
</tr>
</tbody>
</table>

What the examples in bold have in common is that both elements repeat the same concept. It does not really matter whether they are fixed or unfixed in order. What matters is that in such pairs the meaning is repeated, which raises the question: why repeat? In all the other examples, this or some other semantic relationship will be there to aid such a query, together with all the structural considerations (rhythm, alliteration, emphasis, and so on). In the highlighted pairs, the semantic relationships between the constituents are simply irrelevant.

This difference in the semantic nature of binomials has never been pointed out. In my opinion, it introduces a crucial distinction into the purpose of MWUs in standardising legal texts and their status as a fixed unit. As a working label, I would like to suggest *binomials proper* for the coordinated word-pairs with semantic repetition. To determine their frequency in legal texts, the corpus has been searched manually for coordinated pairs which satisfied the definition of a binomial. The material rendered as many as 652 noun tokens, 492 verb tokens, 117 adjectival binomials, and 30 coordinated adverbs.\(^9\) The discussion below has been narrowed down to nouns and verbs, as they are the most prominent grammatical categories.

A puzzling issue to solve, which comes out of the legal material analysed, is that a large proportion of MWUs appear only once in the corpus. It becomes difficult to tell whether such a

---

\(^8\) Valuable as it could prove for this analysis, the recently released *Linguistic Atlas of Older Scots* could not be searched for MWUs by means of automated online queries (http://www.lel.ed.ac.uk/ihd/laoa1/laoa1.html). In order to expand my database in the post doc project, I am going to use the raw text files of the atlas, with the kind consent of Dr Keith Williamson (University of Edinburgh). This will also mean expanding the chronological scope, as the current version of the atlas covers a period up to 1500.

\(^9\) See Table 1 for total word counts.
construction bears the features of a fixed, lexicalised expression. One of the possibilities is to look through later and earlier texts in search of the same strings, to achieve a diachronic range of usage. What one can do synchronically, is to consult dictionaries and context. Unfortunately, authoritative reference tools, such as the Oxford English Dictionary (OED) or the Dictionary of the Scots Language (DSL), do not treat multi-word expressions consistently, and do not help to establish their lexical status. One must thus rely on understanding the surrounding passage and take the best-informed decision, based on the information provided by the dictionaries for individual constituents of a pair, and be sensitive to the context in the document.

Thus, the methodology for this project included the following steps. First, the meaning of individual constituents and the relationship between these meanings has been established with the help of the OED and DSL. Then, bearing in mind Malkiel’s comment about “the fluidity of any semantic classification” (1959: 129) and earlier criticism of any detailed set of semantic relations (e.g., Lambrecht 1984), the following types of semantic relationship in the binomial pairs were established (also drawing on earlier suggestions of semantic categorisation in Cooper and Ross (1975), Koskenniemi (1968), and to some extent Malkiel (1959)):

- **a) binomials proper:** e.g., brother and fraternite, escheit and tak
- **b) complementation:** e.g., form and tenor, heir and se
- **c) cause and effect:** e.g., away takar and hauar, fundin and deliuerit
- **d) contiguity (metonymy):** e.g., gude ground and foundament, mak and performe
- **e) hyponymy:** e.g., infeftment and chartour, dowblit and triblit
- **f) antonymy (contrast):** e.g., byaris and sellaris, persewand and defendand
- **g) metaphor:** e.g., articlis and fundamentis

My categorisation aims to separate the group involving semantic repetition — binomials proper — clearly. The notion of synonymy could perhaps be used here instead, the term however being imprecise and bordering too much on contiguity. I also wanted to stay within the framework developed by Wang (2005), which stresses repetition as a force leading to the fixedness of a given coordinated pair. Finally, I checked the frequency of all category types in the material (Figs. 1 and 2), while the relevant tokens were counted in the analysis of hapaxes (Figs. 3 and 4).

Figure 1 shows that in nouns there are two major types of semantic relation between the constituents of a binomial pair: semantic repetition and complementation. The latter’s high frequency can be easily explained by referring to the characteristics of legal language: complementation as a means to greater precision.

Fig. 1. Frequency of semantic relations in nouns

![Frequency of semantic relations in nouns](image)

---

10 Bauer (1983: 45) allows for the possibility of accepting nonce formations into the linguistic repertoire by the users, on condition that a user is not conscious of prior employment of a specific construction.

11 The DSL includes the Dictionary of the Older Scots Tongue (1100-1700) as well as the Scottish National Dictionary (1700-1970s) together with its 2005 Supplement.

12 No instances of metaphor in verbs were found.

13 The frequency for semantic relations has been normalised per 1000 words.
Still, it is clear that binomials with semantic repetition are the most frequent. What then are the implications for the status of binomials as multi-word units of meaning? It seems that in the cases when the semantic relationship between the constituents introduces new information, such as meaning extension, complementation or contrast, the unit exhibits meaning compositionality and as such cannot be perceived as a lexicalised entity. Of course, the most frequent non-repetitive, complementary expressions, such as *provest and bailies*, or *form and tenor*, are fixed at the pragmatic level (see the earlier discussion on pragmatic non-compositionality) but they will still be transparent at the semantic level. In fact, these two binomials boost the number of complementation pairs in the material, as they constitute almost 60% of all instances.

When it comes to verbs (Fig. 2), the balance is even more strikingly shifted towards binomials proper.

Fig. 2. Frequency of semantic relations in verbs.

Their appearance in the text is not motivated by the need to hone the meaning, make it more precise, extend it, or make it less obscure. They are rather a stylistic device, close to what Koskenniemi postulated for the literary use of binomials. This would run counter to her claim that binomials in legal language are primarily used to enhance understanding and “present a real alternative” (1968: 78). The findings suggest that the creation of binomials was governed by prosodic, etymological or other factors, and not by meaning. Binomials proper are clearly examples of Moon’s pragmatic non-compositionality but also semantic non-compositionality to some extent. In the texts, expressions such as *actis and statutis*, or *statut and ordanit*, are not subject to semantic analysis – both parts mean the same and the whole expression becomes institutionalised, if not lexicalised and fixed. Thus the meaning of a binomials proper is not a sum of A + B, nor some extended meaning of the whole, but rather it is the same even if expressed by one member of the pair only.

Recognising the intricacies of the relationship between frequency and fixedness (Wray 2002: 25, 31–33), the corpus was searched for potential nonce formations. The expectation was that binomials proper would be more formulaic, so that there would be less instances of hapax legomena classifiable as a binomial proper. The material supports this claim (Fig. 3).
In nouns, binomials with semantic repetition are not the most formulaic group, beaten by pairs with complementation involved, but this is due only to the two very frequent lexical pairs mentioned above: provest and bailies, and form and tenor. Still, binomials proper show less variety than other groups – a tendency which is even stronger in verbs (Fig. 4).

4. Conclusions

In analysing the semantic status of multi-word units in legal discourse, it is useful to distinguish between units which are non-compositional, at least pragmatically, and units which are compositional. The latter are not the outcome of a lexicalisation process: they are more diverse in form (see the ratio of hapaxes, Figs. 3 and 4), and allow many different internal semantic relations, which is important for the legal context. Binomials proper, on the other hand, are pragmatically non-compositional, even if the meaning of the constituents can be considered transparent. It is these expressions that are found in the corpus most frequently (Figs. 1 and 2), which suggests that the verbose character of legal discourse, flagged in every related study, may indeed be caused by the employment of binomials proper.
There is another conclusion to be drawn from the present analysis. One has to be ready to accept the fact that in specialised legal discourse, at least at the initial stage of its modern standardisation, the lexicon was selected not only for functional purposes but also for aesthetic, rhetorical and stylistic reasons. It follows from the material analysed that the most formulaic phrases, on the basis of frequency and low type-to-token ratio, were not conditioned by semantic factors. What falls outside the scope of meaning, connected to the utilitarian aspect of legal writings, is the stylistic selection governed by connotation, fashion, sound effects and ready-made cliches. It is therefore justifiable to posit that this is where binomials proper come from and why they can be found in legal texts.

References


Baldwin, Timothy. 2006. “Compositionality and multiword expressions: Six of one, half a dozen of the other?”. Invited talk given at COLING/ACL, Sydney, Australia.


