

John Benjamins Publishing Company



This is a contribution from *English Historical Linguistics 2008*.

Selected papers from the fifteenth International Conference on English Historical Linguistics (ICEHL 15), Munich, 24-30 August 2008. Volume II: Words, texts and genres.

Edited by Hans Sauer and Gaby Waxenberger.

© 2012. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible to members (students and staff) only of the author's/s' institute, it is not permitted to post this PDF on the open internet.

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

Please contact rights@benjamins.nl or consult our website: www.benjamins.com

Tables of Contents, abstracts and guidelines are available at www.benjamins.com

Repetitive and therefore fixed?

Lematic bundles and text-type standardisation in 15th-century administrative Scots

Joanna Kopaczyk

Adam Mickiewicz University, Poznań

This is a pilot study investigating the role of phrasal fixedness in the development of a standardised text type. The linguistic material comes from the Edinburgh Corpus of Older Scots (ECOS), consisting of samples of administrative records from 15th-century Scotland. The corpus has been searched for re-occurring lemmatic bundles, which are the indicators of emerging patterns and standardising usage in the records, developing in the context of linguistic standardisation of Scots. The findings are interpreted with regard to their semantics and function in the records, and indicate that the text type as such was not yet fully standardised in its repertoire of fixed phrases serving a specific purpose. In individual locations, however, one finds a greater degree of consistency and a tendency to develop a local norm. Similarly, in some specific textual functions the lexical fixedness may be present to a larger extent than in others.¹

1. Text-type standardisation

Standardisation affects all levels of language. This is because variation can be found at all linguistic levels and may counteract with the communicative needs of the users. Out of a pool of possibilities, a gradual selection of most efficient forms takes place and its results disseminate among the language community. In other

1. This pilot study was initially presented at the 15th International Conference on English Historical Linguistics at Ludwig-Maximilians-Universität München in 2008. I am grateful to the conference participants for their inspiring comments, which guided me through the initial stages of the project. Since then, a larger-scale version of the study has attracted the financial support of the Polish Ministry of Science and Higher Education (grant nr N N104 014337), and acquired its final shape in a monograph (Kopaczyk forthcoming c) and several smaller publications (Kopaczyk 2011, forthcoming a, forthcoming b).

words, language users converge on a developing norm.² Whatever the level of linguistic analysis, standardisation always involves the notions of correctness and appropriateness. Only when referring to standard usage can one say that a given linguistic choice in somebody's performance is incorrect or inappropriate. Usually in historical studies priority is given to spelling standardisation and grammar standardisation. In this paper, however, I concentrate on the level of *text*, which in my view is also subject to standardisation. Language users, both encoders and decoders, have certain expectations as to the form of a text serving a specific function. We learn the standard versions of text types while we acquire our linguistic competence. This is why the content and structure of a cookery recipe is relatively obvious to everybody, just as we know what vocabulary or fixed phrases to expect in a letter of complaint or an obituary. These forms have developed in time, and thus can be approached diachronically.

The present investigation concentrates on the degree of standardisation and fixing of legal administrative written conventions, on the basis of local documents written in a standardising vernacular, Scots. 15th and early 16th-century Scots was a language on its way to standardisation (Devitt 1989; Bugaj 2004), slowly converging on a specific selection of linguistic variants accepted and used throughout the country. The language of administration and law was also becoming more and more fixed in its idiom. This study has a qualitative nature and aims to show that throughout Scotland notaries and local scribes were very much aware of the appropriateness of their selection of constructions. This point will be illustrated by the most frequent formulaic expressions occurring in the corpus material.³

2. Investigating the link between repetition and fixedness

The existence and importance of fixed expressions in language has long been recognised. Jespersen, for instance, noticed that “a language would be a difficult thing to handle if its speakers had the burden imposed on them of remembering every little thing separately” (1976 [1924]:85). We speak and write using

2. For a seminal discussion of standardisation as a diachronic process, taking place in inter-related stages, see Haugen (1966); for a cultural interpretation of linguistic standardisation see Geeraerts (2003); a diachronic look on the standardisation of English, also in various text types, can be found in an important collection of papers edited by Wright (2000).

3. This is a qualitative pilot study for my post-doctoral project and at this stage statistical counts are not possible. For a full discussion of the final research results of the project, see Kopaczyk (forthcoming c).

our favourite chunks, whether learnt as appropriate for the occasion, or simply designed by ourselves more or less consciously. This important feature of usage helps forensic linguists carry out idiolect studies. The preference for uniformity and stability of linguistic expression, which is sanctioned by tradition and institutionalisation, is typical for specialised discourse, most importantly for legal language.

With this in mind, a question can be asked: how such collocations or co-occurring strings are or have been formed? So far, research into phrasal fixedness has been largely synchronic in nature, and focussed on language acquisition, either first or second, and the processing of multi-words units in the brain (see the reference list in Wray 2002). Fixedness, according to Hudson (1998: 1), is a process which turns one-off combinations of words into “fixed expressions that become units in their own right”. Although potential links to lexicalisation and grammaticalisation are visible here, the present paper will address the notion of fixedness on the level of text type. The idea is to perceive repetition and fixedness as forces leading to linguistic standardisation. Repetition in legal administrative texts can be noticed on several levels, as confirmed in many analyses of legal language (Mellinkoff 1963; Goodrich 1990; Tiersma 1999; Gibbons 2004; Williams 2005): structural parallelism in sentences and phrases, lexical repetition forming cohesive networks in the text, a significant degree of constituent coordination (e.g. in repetitive word pairs), semantic repetition in the so-called binomials proper (Wang 2005; Kopaczyk 2009). The working hypothesis is that repetition leads to fixing and standardising of a text type, fossilising certain expressions and making the text type more formulaic.

3. Extracting fixed lexical strings and their functional interpretation

The present line of inquiry was inspired by my research into binomial expressions (Bugaj & Włodarczyk 2006; Kopaczyk 2009). When searching Scottish and English burgh records for binomial pairs, I noticed a significant number of repetitive expressions larger than a coordinated pair. Quite often these were, as if, expanded binomials, as in *landis gudis possessiounis and vther thingis*, but there were also strings without coordination, as in *of a good man of the town*. Such expressions, frequent and fixed in structure as they were, proved to be difficult to extract systematically by a regular search query as they did not have to include any common element (such as a coordinating device, as in the case of binomials or even multinomials).

In language acquisition research, a number of methods for extracting formulaic sequences have been suggested. Of course, one can establish such chunks intuitively

or on the basis of shared knowledge (Wray 2002: 20–25), bearing in mind the problems inherent in such procedures. Indeed, the relationship between repetition and fixedness of a text-type in historical specialised discourse is best investigated by means of corpus methods. As Biber et al. remarked, “historical corpora now offer the opportunity to investigate the use of many linguistic features across historical periods or to examine the development of *registers* [emphasis mine] over time” (1998: 12). The current project draws upon Biber’s text-type research methodology (1989 and later publications), which involves looking for co-occurring linguistic features in order to see what creates a given register. More specifically, I adapted the method of extracting the *lexical bundles* (Biber et al. 1999), which are strings of lexical items repeated frequently in an unchanged form in a given type of text. The adaptation of this method consisted in extracting *lemmatic bundles*, see Section 4. for the explanation of the rationale and the procedure. Still, in each approach one has to decide about the length of the searched strings, and what frequency will actually count as significant.

These two parameters – length and frequency – will depend on the shape and extent of the corpus but, in any case, such choices may come across as subjective and arbitrary (see Kopaczyk forthcoming b for an overview of cut-off points in lexical bundle research). Wray (2002) sees further problems in computer searches as they may produce sequences which in context are not formulaic, yet resemble those which may be intuitively interpreted as such, for instance “*keep your hair on* is not formulaic when it means ‘don’t remove your wig’” (Wray 2002: 31), as opposed to ‘calm down’ when it is formulaic. In the case of the method applied in the present project, this important remark has little bearing, though. The texts in question are homogeneous in their purpose and subject matter and such coincidences of alternative meanings are unlikely to be found. Nevertheless, frequency has been confirmed in other studies to be a valid criterion in the search for formulaic sequences, be it in written or spoken discourse (DeCock et al. 1998; Clear 1993; Butler 1997; Moon 1998a; Moon 1998b).

The corpus of texts used for this study (see Section 4 below) shares situational characteristics, and thus is written in the same register. In Biber’s terms, registers are instances of linguistic usage “defined in terms of a particular constellation of situational characteristics” (2001: 3). In reference literature one comes across a terminological maze in which registers, genres and text types are used in an overlapping fashion. To keep things simple, I would like to use the notion of a *text type* as referring to a combination of similar linguistic choices, whereas *register* will refer to external forces shaping the text type. This approach evokes Hallidayan functional distinctions, based on non-linguistic parameters of a communicative situation, such as the *field* (what is the context and the purpose of the communicative act?), *tenor* (what is the relationship between participants?) and *mode*

(how is the communicative act performed?) (Halliday 1979). In the case of legal writing, and administrative records in particular, the non-linguistic parameters are stable regardless of the provenance of a given text. The texts are written in a formal setting to inform about legal actions and laws introduced (field), the author takes the floor in a mono-directional fashion, writing from the position of authority but making reference to other participants in the legal action (tenor), and the mode of communication is written. What is interesting, is to see how such administrative texts were written in terms of lexico-grammatical choices, and whether any emerging patterns contributed to the development of a standardised text type. The fixed lexical strings recurrent in the corpus have been categorised according to their meaning and function, roughly following the Hallidayan approach outlined above.

4. The corpus and the method

The *Edinburgh Corpus of Older Scots* (ECOS)⁴ was compiled as the database for the *Linguistic Atlas of Older Scots* (LAOS), which can be searched online for historical dialectal features of Scots grammar and lexicon. The database is a representative collection of text samples from almost every county in Scotland and amounts to c.390,000 words. These texts are uniform in terms of register because they all come from the area of law and administration. They are local documents of various kinds, localisable on the basis of extralinguistic information. The full list of locations referred to in the present pilot study may be found in the Appendix together with the codes used in the discussion.

The time span of the corpus is about 120 years, from c1379 to 1500, with a slant towards the later part of the 15th century. In the present study, each packet of texts from each county has been treated as a unit.⁵ I also regarded all documents included in the corpus as instances of administrative activity and did not introduce further text type categorisations. However, if need be, it should be possible in future research.

To implement the lexical bundles method (Biber et al. 1999), one needs to run the corpus through software which automatically extracts repeated lexical strings of a given length. A major problem with applying this method to historical

4. I would like to thank Dr Keith Williamson from the University of Edinburgh for giving me access to the files.

5. In my post-doctoral dissertation I introduce finer chronological distinctions (see Kopaczyk forthcoming c).

corpora is posed by spelling variation, whereby spellings such as *other*, *oper*, *uthir* and *vthir* will constitute parts of different lexical bundles.⁶ The way to overcome this difficulty was to utilise the lemmatic tags attached to each word form in the corpus. Thus, all the examples above would be listed in the lemmatised ECOS files as OTHER. The procedure was to extract the actual lemmata which formed longer co-occurring strings. I ran the tagged version of the ECOS through a programme which looked for co-occurring bundles of adjacent lemmata.⁷ The strings I extracted may therefore be called *lemmatic bundles*. In this manner the problem of spelling variation in the corpus was largely avoided, however the results prove that the success of this method will depend on the consistency of lemmatisation.

By definition, the lemmata serving as tags for specific word forms are stripped of any morphological marking. This is why lemmatic bundles come across as ungrammatical (see examples in Section 5). I solved this problem by translating the lemmatic bundles according to the context provided by the actual text. In other words, each lemmatic bundle was cross-checked against its manuscript context, and an appropriate translation was entered into the tabulated data. Another point to bear in mind while interpreting the data is that the lemmata are written in English or in standardised Scots spelling, which again does not capture the reality of the manuscripts written in 15th-century Scots. The features of spelling or underlying phonology are not in focus here, however one should be aware of the fact that lemmatisation obscures some typically Scots lexical items, giving them English or modern Scots labels.

For each county, a list of co-occurring strings of three to eight elements has been extracted. The expressions were arranged from most to least frequent, and cut off at ten occurrences per text.⁸ As there were numerous examples of fixed chunks occurring less than ten times, the above methodological decisions allowed for a manageable qualitative analysis. I was also interested to see which length

6. This pilot study predates Culpeper and Kytö's (2010) work on early modern English dialogues, where they also extracted most frequently repeated lexical strings from a historical corpus. In order to do that, they unified the spelling automatically by means of special software, VARD (Baron 2012). I partly make use of this software in the final version of the project (Kopaczyk forthcoming c) but at the pilot stage I had to look for a different solution to the spelling variation problem.

7. I would like to thank Mr Dariusz Stróżyński (BSc) for designing the software.

8. Researchers vary when it comes to establishing the formulaicity thresholds: Altenberg (1998:101) treats any lexical string appearing more than once in the same form as formulaic, while other studies set the cut-off point at ten occurrences (Butler 1997:66) (for a further discussion of cut-off points, see Kopaczyk forthcoming b). In view of such discrepancies, I decided to make a data-driven decision and establish the pilot threshold at ten instances.

setting would render the most fixed and frequent examples, in view of Biber et al.'s (1999) observation that repetitive strings of three elements make up around 25% of the word count in the corpus of modern English conversation, and 18% of the word count in the academic prose corpus. Compared with that, repetitive strings of four elements accounted for three and two per cent of word count respectively (1999: 993–994). To assess the presence of fixed strings of different lengths in the pool of the most frequent bundles, I conflated the findings for different lengths of lemmatic bundles. The results point to the presence of longer strings among the most frequently repeated elements (see Section 5).

The final methodological decision was to select only those strings from the extracted bundles which have a complete phrasal structure. This decision was due to the preliminary character of the study and was driven by the need to establish whether the automatic extraction method would, indeed, render complete grammatical structures. Interestingly, however, large quantities of recurrent strings consisted of a part of a phrase, or stretched across phrasal boundaries, suggesting even longer fixed passages. I categorised and discussed the non-phrasal bundles in Kopaczyk (forthcoming c). Of the phrases in the final result list in the pilot study, all were connected to the purpose of the text. There were no random structures which would have no bearing on the type of text.

5. Semantic and functional categorisation of lemmatic bundles

All lemmatic bundles extracted from the corpus were sorted by frequency of occurrence. Then the top one hundred were arranged into tables, following a semanto-functional categorisation outlined in Section 3. The examples in the tables are numbered consecutively (Column 1). Bundles are listed in Column 2 in their lemmatic shape and are provided with a modern English contextual translation in Column 3. The number of occurrences is given in Column 4. The last column in every table includes information on how many locations utilise a given repetitive string. The codes used for individual counties are expanded in the Appendix.

5.1 Expressions relating to the field: External reality

Administrative records were drawn in a certain situational context. In the material, there are fixed expressions pertaining to various aspects of this context, for example time and space or the objects of legal action.

When it comes to temporal reference in the documents (Table 1), the most formulaic bundles refer to the day and the month of the proceedings. The bundles do not reveal which month or which actual date was most frequently mentioned,

but they provide a kind of a stable frame which was completed in the texts with different information depending on the time of the entry. It is worth noticing that a potential reference to weeks or some other temporal divisions is missing from the most formulaic strings. It is also interesting that the year is attributed to divinity, ‘the year of God’ (3), which constitutes the most typical reference to the year in four counties.

Table 1. Temporal reference

#	Lemmatic bundle	Translation	Count	Counties
1.	of y ^e month	of the month	143	FIF, PBL
2.	day of y ^e month	day of the month	125	FIF, PBL
3.	ye year of god	the year of God	108	ABD, AGS, FIF, PTH
4.	of ye month	of the month	96	ABD, AGS, FIF, PTH
5.	y ^e same day	the same day	92	FIF
6.	day of ye month	day of the month	90	ABD, AGS, FIF, PTH
7.	time to come	time to come	62	ABD, AGS, EDB, FIF
8.	& then incontinent	and then immediately	62	FIF, PBL

The bundle method reveals standardised usage. For instance, example (5) is a string repeated ninety two times in one location in an unchanged form. This is possibly an indication of a locally prevalent fixed construction. Other counties must have rendered the same meaning in a less formulaic fashion, or perhaps this specific aspect of meaning was not present elsewhere to such an extent. The phrase ‘the same day’ serves two purposes: on the one hand it makes a temporal reference, while on the other hand it sustains discourse cohesion by calling upon a previously given date. The same tendency to fix textual cohesion devices will be illustrated in 5.3 below. The remaining bundles in (7) and (8) indicate formulaic future reference and a sequence of events, respectively.

There is one more technical observation to be made at this junction, namely that examples (1) and (4) and examples (2) and (6) are listed separately, in fact, only because of the spelling of one lemma. The form y^e ‘the’ represents a thorn with a superscript ⟨e⟩, which the corpus compiler decided to distinguish from ye ‘the’ without the superscript in the list of corpus lemmata. This inconsistency revealed by the present pilot study has inspired a change in the spelling unification methodology in the final version of the project (Kopaczyk forthcoming c). This technical deficiency notwithstanding, the standardising chunks of text are visible in the lemmatic bundles extracted from ECOS in a more consistent manner than they would be, had they been extracted straight from the corpus by means of a

typical lexical bundle methodology. Of course, spelling variation is a problem to be borne in mind and overcome in any automated corpus searches, which was one of the major challenges for the later stages of this project.

Table 2 presents another aspect of extralinguistic reference: the reference to the place. As the documents comprised in the corpus often pertain to land transfer, the most frequent bundles in this section, examples (9) and (10), are indications for land measurements, which would introduce specific values in the actual texts (cf. the date reference above). Similarly, example (21) specifies the direction of a given measurement.

Table 2. Place reference

#	Lemmatic bundle	Translation	Count	Counties
9.	on one part	on one part	152	ABD, AGS, AYR, FIF, LNK, PBL, PTH, STG, XLOC
10.	on other part	on other part	140	ABD, AGS, AYR, FIF, LNK, PBL, STG, XLOC
11.	y [^] e say burgh	the said burgh	76	FIF
12.	of y [^] e town	of the town	56	AYR, FIF
13.	in plain court	in open court	54	ABD, AYR, FIF
14.	to yis court	to this court	50	LNK
15.	y [^] e burgh court	the burgh court	47	FIF
16.	my lord court	my lord's court	45	LNK
17.	in y [^] e chapel	in the chapel	45	FIF
18.	ye next court	the next court	44	FIF, LNK
19.	of my lord court	of my lord's court	43	LNK
20.	of ye town	of the town	43	FIF, LAWS
21.	on y [^] e west	on the west ⁹	42	PBL

The majority of formulaic bundles in this category refers to the actual court (examples 12–16, 18–19) or another place where the court could be held, such as the chapel (example 17). The more general reference to the burgh or town, where

9. Interestingly, the reference to the eastern direction, *on y[^]e east*, is very close in terms of frequency, with 37 instances in the corpus. It fell, however, outside the predetermined number of most frequent examples.

a particular document was compiled, is also present among the repetitive strings, as in (11), (12) and (20). The bundles reveal the most typical modification of these nouns. The notary in Fife comes up with a compound *burgh court* in (15), a phrase which was not used with such frequency elsewhere. Other modifiers describe the court as *plain* ‘open, public’ (example 13) or as belonging to the lord – in this case, the king¹⁰ (examples 16 and 19). They also place the court in the temporal context, as in (18), which indicates a frequent forward outlook, *ye next court*, rather than reference to some past events.

Tables 3a and 3b present the most frequent lemmatic bundles concerning legal activity, either with focus on the dealings of the court, or with focus on the actions of the people involved. The bundles have been subdivided only for the purpose of a more convenient discussion, because in practice both perspectives overlap.

Table 3a. Legal action: Focus on the court

#	Lematic bundle	Translation	Count	Counties
22.	give for doom	given for judgement	124	AYR, FIF
23.	suit call y ^e court affirm	suits called the court affirmed	84	FIF, PBL
24.	be give for doom	was given for judgement	83	FIF
25.	y [^] t be give	that was given	77	FIF
26.	call y ^e court	called the court	72	FIF, PBL
27.	y ^e court affirm	the court affirmed	68	FIF, PBL
28.	& y [^] t be give	and that was given	68	FIF
29.	& no [^] t compear	and did not appear	64	LNK
30.	y ^e suit call	the suits called	61	FIF, PBL
31.	call & no [^] t compear	called and did not appear	57	LNK
32.	y [^] t be give for doom	that was given for judgement	54	FIF
33.	y ^e suit call y ^e court affirm	the suits called the court affirmed	50	FIF, PBL
34.	& y [^] t be give for doom	and that was given for judgement	46	FIF
35.	ofttimes call & no [^] t compear	many times called and did not appear	45	LNK
36.	ye suit call	the suits called	43	FIF
37.	ye court affirm	the court affirmed	42	FIF

10. The texts comprising the ECOS corpus were compiled in the royal burghs. For an overview of different kinds of burgh ownership in Scotland, see, e.g. Pryde (1965).

When it comes to the reference to legal activity, several burghs display a larger degree of formulaicity than others. The extracted bundles come mostly from Fife, Peebles and Lanark. The notary responsible for the documents in the first location had a formulaic way of closing a given passage: “and that was given for doom” (34). This formula seems unique for that particular author and he used it 46 times in the corpus. Moreover, the bundle method reveals the formulaic nature of elements which create this six-element string, as visible in examples (22), (24), (25), (28) and (32).¹¹ They were more frequent than the 6-gram and could be surrounded with other contextual possibilities. For example, the shortest bundle, the 3-gram “given for doom”, was encountered in the corpus over 120 times, which opens the possibility of different contexts introducing this bundle than “and that was”. Changing “that” to “this” or “which” would render a different bundle, not frequent enough to reach the top one hundred bundles.

Another group of overlapping bundles pertaining to the legal activity in court can be comprised into the string “the suits (which were) called, the court affirmed”, or “the suits (were) called, (which) the court affirmed”, see examples (23), (26), (27), (30), (33), (36) and (37). This formula was mostly used by the scribes in Fife and Peebles. The final formulaic string in this category is related to the fault of non-appearance in court (examples 29, 31, 35), which must have been a notorious problem in Lanark, or so the records indicate. Examples (47) and (48) in Table 3b refer to the same situation. It is plausible that other locations faced a similar difficulty in compelling witnesses and other participants to appear in court, but the relevant passages in the records must have been phrased in a more diversified manner.

The participants in legal proceedings engaged in a range of activities, some of which were given a formulaic reference in the records. They could witness the proceedings (examples 38 and 41), be obliged to follow some resolutions (examples 40, 42, 43, 45), and answer to challenges (example 46). Most of the fixed strings related to these activities appeared in records from several counties, pointing towards a uniformity of usage across the burghs. Another interesting formula emerging in the bundle extraction was the reference to becoming a fully recognized citizen of a burgh, see example (44).

11. I used the term *paradigmatic overlap* to refer to the situations where shorter bundles are included in longer bundles, as opposed to *syntagmatic overlaps*, where only a part of a given bundle is shared with another formulaic string (Kopaczyk forthcoming c).

Table 3b. Legal action: Focus on the people

#	Lematic bundle	Translation	Count	Counties
38.	witness of the-which-thing	witness of the which thing	119	AGS, BWK, EDB, ELO, FIF, PTH, XLOC
39.	be in amerciament	is under penalty	119	FIF, LNK
40.	bind & oblige	bind and oblige	95	AGS, AYR, PTH, STG, XLOC
41.	in witness of the-which-thing	in witness of the which thing	70	AGS, FIF, PTH, XLOC
42.	be to remember	is to remember	66	ABD, PBL
43.	be it ken	be it known	59	AGS, FIF, PTH
44.	be make burgess	was made burgess	57	PBL, FIF
45.	jt be to remember	it is to remember	46	PBL
46.	to ye challenge	to the challenge	42	AYR
47.	fault of entry	failure to appear (in court)	42	LNK
48.	for fault of entry	for the failure to appear (in court)	42	LNK

In Table 4, the bundles pertaining to the objects of legal action have been gathered. As mentioned above, one of the most frequent topics in the proceedings was land transfer, which also finds support in the most numerous bundle in this category, *the said land* (examples 49, 53 and 55).

Table 4. Objects of legal action

#	Lematic bundle	Translation	Count	Counties
49.	ye say land	the said land	190	ABD, AGS, AYR, EDB, FIF, PTH, ROX, STG, WLO, XLOC
50.	earth & stone	earth and stone	126	AYR, FIF, PBL
51.	yir present letter	this present letter	106	ABD, FIF, PTH, XLOC
52.	fraud or guile	fraud or deceit	87	ABD, AYR, EDB, FIF, PTH
53.	y [^] e say land	the said land	70	AGS, PBL, PTH
54.	by yir present letter	by this present letter	66	FIF, PTH, XLOC
55.	of ye say land	of the said land	51	AGS, EDB, FIF, XLOC
56.	of usual money	of usual money	49	FIF, PTH, XLOC
57.	in y [^] e hand	in the hand	49	AYR, PBL

Other bundles refer to objects brought to court, e.g. *earth and stone* (50), which was symbolically used to represent the actual plot of land under transaction. Possibly, the bundle in (57) makes reference to the way of receiving such tokens. A more abstract object of legal activity has emerged in the form of a frequent offence, *fraud or deceit*, in bundle (52). The writs announcing regulations and legal decisions were also referred to in a formulaic manner, see examples (51) and (54). Another emerging formulaic string is connected with the financial dealings in the local currency (example 56).

5.2 Expressions relating to the tenor: Author reality, participant reality

When it comes to the labels for the participants of the communicative situation, the reference becomes more formulaic. More phrasal bundles are repeated in a variety of locations, and the token numbers are higher. For the purpose of a clearer data presentation, I have separated the bundles pertaining to the hereditary relations (Table 5a), to the authorities (Table 5b), and listed the other relevant bundles in the last table (Table 5c).

Table 5a. Participants: Hereditary relations

#	Lematic bundle	Translation	Count	Counties
58.	heir or assignee	heir or assignee	110	ABD, AGS, AYR, EDB, FIF, STG, XLOC
59.	heir & assignee	heir and assignee	96	ABD, AYR, EDB, PTH, STG
60.	& his heir	and his heir	81	AGS, AYR, FIF, PBL, XLOC
61.	executor & assignee	executor and assignee	61	AGS, FIF, NRN, PTH
62.	heir executor & assignee	heir, executor and assignee	59	AGS, NRN, FIF, PTH
63.	his heir & assignee	his heir and assignee	42	EDB, PTH, STG

In Table 5a, the issue of heirship comes to the fore in multiple locations. What is interesting is that the formulaic chunks of discourse related to hereditary relations have the form of binomials and multinomials (cf. examples 40, 50 and 52). Similarly, several bundles from Tables 5b and 5c also have the same formal characteristics (examples 66, 74 and 78), which indicates the disposition of legal discourse towards coordinated constructions (for a more detailed discussion of different semantic relations between the coordinated elements, see Kopaczyk 2009).

Formulaic reference to authority is among the most frequent discursal strategies in the corpus, with more than two hundred tokens of the top bundles

(examples 64, 65). It is the lexeme *lord* that seems more conducive to formulaic behaviour, though. Other variant bundles related to *lord* are listed under (67), (68), (70), (71) and (73).

Table 5b. Participants: Authorities

#	Lemmatic bundle	Translation	Count	Counties
64.	ye say lord	the said lord	234	ABD, AGS, DNB, EDB, FIF, LNK, PTH, XLOC
65.	ye say sir	the said sir	202	AGS, AYR, EDB, ELO, FIF, PTH, XLOC
66.	abbot & convent	abbot and convent	89	AGS, FIF, ROX
67.	y ^e say lord	the said lord	59	LNK, PTH, XLOC
68.	of my lord	of my lord	58	LNK
69.	ye say abbot	the said abbot	56	AGS, FIF, ROX
70.	of our lord	of our lord	52	AGS, FIF
71.	to ye say lord	to the said lord	51	ABD, EDB, FIF, XLOC
72.	ye say earl	the said earl	50	DNB, PTH, STG
73.	my lord follow	my lord followed	45	LNK
74.	ye say abbot & convent	the said abbot and convent	43	AGS, FIF, ROX

Table 5c. Other participants

#	Lemmatic bundle	Translation	Count	Counties
75.	til all man	to all men	110	ABD, AGS, AYR, FIF, PTH, XLOC
76.	before yir witness	before these witnesses	110	AYR, FIF, PBL, PTH
77.	ken til all	known to all	94	ABD, AGS, FIF, PTH, XLOC
78.	all & sundry	all and sundry	93	ABD, AGS, BWK, EDB, FIF, PBL, XLOC
79.	ken til all man	known to all men	92	ABD, AGS, FIF, PTH, XLOC
80.	be it ken til all	be it known to all	51	AGS, FIF, PTH
81.	be it ken til all man	be it known to all men	51	AGS, FIF, PTH
82.	y ^e say bailie	the said bailie	46	PBL

In Dunbar, Perth and Stirling, there is also a formulaic reference to *the said earl* (72), which could, in fact, be rendered with *the said lord*, but the scribes in these locations chose to be more specific. Finally, in Table 5b I also included the

bundles connected with the religious context, such as (66) and (74). Without a larger context it is difficult to say, though, in what capacity the abbot and convent appeared in the documents: the petitioners, the owners of goods or land, or perhaps the offenders.

The formulaic string of major importance in this group is the announcing formula “be it known to all (men)”, which was used in at least three counties (examples 80 and 81). Its shorter formulaic ingredients enjoyed a wider popularity (examples 75, 77 and 79). This formula could be interpreted as a directive to the general audience of the document, including “all men”, so all the inhabitants of a given burgh. The fact that legal and administrative texts aimed to be comprehensive in their bearing, is also signalled by the formulaic bundle *all & sundry* (78). This expression was frequently employed in a wide selection of locations to indicate that the documents were aimed at the whole population. Finally, example (82) refers to a citizen of a burgh, or, more precisely, a representative of the municipal authorities, who is involved in legal proceedings. The pronominal modification with *the said* contributes to the cohesion of the text, making reference to the participant who has already been mentioned. More bundles performing a cohesive function are listed in Table 6a below.

5.3 Expressions relating to the mode: Channel reality

One of the most interesting outcomes of the lemmatic bundle search is the collection of fixed phrases used throughout the corpus for the sake of textual cohesion (Table 6a) and narrative sequencing (Table 6b).

The lemmatic bundle in (83) is, in fact, the most frequent fixed expression in the corpus. It is present in a wide range of counties and functions as a general cohesive device, making reference to previous discourse. Typically, *of the said* would require a nominal head, e.g. *of the said land*, however I have decided to treat this sequence and other similar structures as potential phrases,¹² to be able to include them in this discussion and show their extensive presence in the corpus. The records contain several prepositional phrases used to the effect of creating links within the texts (examples 83–89 and 91–93), so that the content of the administrative document remains internally coherent.

The remaining examples point to standardising narrative strategies, with sequential links (examples 94, 100) and simple action and state verbs (examples 96, 97, 99). It has to be mentioned, though, that the collection of texts in which

12. The prepositional phrase includes an elliptical *the said*, serving as a substantivised adjective, as in “We cater for our clients and answer every question of the said”, which may be stylistically deficient, but theoretically it is grammatical.

the former bundles emerged has been deemed by the corpus compiler as literary. This could be the reason why the sequential links we so consistently used by the author.

Table 6a. Textual cohesion

#	Lematic bundle	Translation	Count	Counties
83.	of ye say	of the said	935	ABD, AGS, AYR, BWK, DNB, EDB, ELO, FIF, LNK, MLO, MRY, NRN, PTH, ROX, PBL, STG, WLO, XLOC
84.	to ye say	to the said	560	ABD, AGS, AYR, BWK, DNB, EDB, ELO, FIF, LNK, NRN, PTH, STG, WLO, XLOC
85.	of y ^e say	of the said	273	AGS, AYR, FIF, LNK, PBL, PTH, STG, XLOC
86.	to y ^e say	to the said	141	AGS, AYR, FIF, PBL, PTH, XLOC
87.	of the say	of the said	123	ABD, AGS, EDB, XLOC
88.	of ye same	of the same	74	AGS, FIF, LNK, STG, XLOC
89.	by ye say	by the said	63	ABD, AGS, FIF, LNK, PTH
90.	as say be	as said is	58	AGS, EDB, STG, XLOC
91.	within y ^e say	within the said	49	FIF
92.	to ye foresaid	to the foresaid	48	ABD, FIF, PTH
93.	to the say	to the said	46	ABD

Table 6b. Narrative bundles

#	Lematic bundle	Translation	Count	Counties
94.	item he say	moreover he said	190	LITX
95.	y [^] t be to say	that is to say	132	ABD, AGS, FIF, PBL, PTH
96.	y [^] t be give	that was given	77	FIF
97.	& y [^] t be give	and that was given	68	FIF
98.	yat be to say	that is to say	67	ABD, FIF, LAWS, PTH
99.	that he have	that he had	63	ABD, FIF, LITX, LNK
100.	item he say that	moreover he said that	54	LITX

Finally, two bundles serve as introductory formulas for some explanation or a more precise rendition of a previous statement (examples 95 and 98), which fulfils the demand that legal discourse be clear and unambiguous.

6. Conclusions

The lemmatic bundle search objectively extracted numerous examples of fixed strings of lexical items, repeated frequently in the records. The findings are indicative of what lexical strings were fixed in which collection of texts, what kind of wording was most popular and which counties were making identical or similar lexical choices (see the county codes in the Tables).

The collection of bundles suggests that the authors of local administrative texts did not use ready templates in the 15th century Scotland. However, certain tendencies in developing stable textual patterns can be discerned. There is a noticeable tendency towards formulaic reference with regard to the participants involved in the communicative act. Also, due to the similarity of topics and cases dealt with, the texts display a similar selection of terms and patterns, only slightly varied grammatically or collocationally. It is quite striking, however, how much formulaicity is encountered in individual locations. This fact is a confirmation of the characteristics of legal language, recalled at the beginning of this paper. Especially in texts from Fife, Lanark and Peebles one can notice a large degree of consistency. It would be interesting to see if Edinburgh norms began to influence document compilation elsewhere later on, or whether such local traditions were difficult to alter and collapsed only after the completion of anglicisation in Scotland.

This line of research, involving objectively extracted fixed lexical strings, seems promising for the discovery of standardising patterns in specialised discourse. The extracted patterns contribute to the communicative functions of the text. They can be classified according to the framework of field, tenor and mode, where each of these functional plains stimulates a preference for different structures. The major methodological difficulty in the automatic identification of identical strings lies in the spelling variation and in the wealth of non-phrasal material which gets extracted together with phrasal constituents. The first problem can be overcome by lemmatising the corpus (taking care to maintain consistency in the choice of the lemmata) and extracting lemmatic bundles, or by artificial spelling unification. The second observation opens a new window onto linguistic fixedness, pointing towards the fact that what gets fixed and standardised does not have to answer to the criteria of structural completeness. Therefore, the next step in this project should be to see how much fixing happens within phrases or across their boundaries, and what non-phrasal bundles emerge from the corpus (see Kopaczyk forthcoming c).

Sources

- LAOS = *Linguistic Atlas of Older Scots*. 2008. (<http://www.lel.ed.ac.uk/ihd/laos1/laos1.html>) [Last access February 2012]
- Williamson, Keith, ed. & comp. 2008. *The Edinburgh Corpus of Older Scots (part 1: 1380–1500)*. University of Edinburgh.

References

- Altenberg, Bengt. 1998. "On the Phraseology of Spoken English: the Evidence of Recurrent Word-combinations". *Phraseology: Theory, Analysis and Applications* ed. by Anthony Cowie, 101–122. Oxford: Clarendon Press.
- Baron, Alistair. *VARD 2*. (<http://www.comp.lancs.ac.uk/~barona/var2/>) [Last access: January 2012]
- Biber, Douglas. 1989. *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber, Douglas, Susan Conrad & Randi Reppen. 1998. *Corpus Linguistics: Investigating Structure and Use*. Cambridge: Cambridge University Press.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.
- Biber, Douglas & Susan Conrad, eds. 2001. "Introduction: Multi-dimensional Analysis and the Study of Register Variation". *Variation in English: Multi-dimensional Studies*, 3–12. London: Longman.
- Bugaj [Kopaczyk], Joanna. 2004. "Middle Scots As an Emerging Standard and Why It Did Not Make It", *Scottish Language* 23.19–34.
- Bugaj, [Kopaczyk] Joanna & Matylda Włodarczyk. 2006. "At 'the wylle and plesur' of 'kyng and souereaine': In Search of Motivations for Binomials As Markers of Legal Language". *IFAtuation: A Life in IFA. A Festschrift for Professor Jacek Fisiak on the Occasion of His 70th Birthday* ed. by Katarzyna Dziubalska-Kołodziejczyk, 85–98. Poznań: Wydawnictwo Naukowe UAM.
- Butler, Christopher S. 1997. "Repeated Word Combinations in Spoken and Written Text: Some Implications for Functional Grammar". *A Fund of Ideas: Recent Developments in Functional Grammar* ed. by Christopher S. Butler, John H. Connolly, Richard A. Gatward & Roel M. Vismans, 60–77. Amsterdam: IFOTT.
- Clear, Jeremy. 1993. "From Firth Principles: Computational Tools for the Study of Collocation". *Text and Technology: In Honour of John Sinclair* ed. by Mona Baker, Gill Francis & Elena Tognini-Bonelli, 271–292. Amsterdam: John Benjamins.
- Conrad, Susan & Douglas Biber, eds. 2001. *Variation in English: Multi-dimensional Studies*. London: Longman.
- Culpeper, Jonathan & Merja Kytö. 2010. *Early Modern English Dialogues: Spoken Interaction As Writing*. Cambridge: Cambridge University Press.
- DeCock, Sylvie, Sylviane Granger, Geoffrey Leech & Tony McEnery. 1998. "An Automated Approach to the Phrasicon of EFL Learners". *Learner English on Computer* ed. by Sylviane Granger, 67–79. London: Longman.
- Devitt, Amy. 1989. *Standardising Written English. Diffusion in the Case of Scotland 1520–1659*. Cambridge: Cambridge University Press.

- Geeraerts, Dirk. 2003. "Cultural Models of Linguistic Standardisation". *Cognitive Models in Language and Thought: Ideology, Metaphors and Meanings* ed. by René Dirven, Frank Roslyn & Martin Pütz, 25–68. Berlin, New York: Mouton de Gruyter.
- Gibbons, John. 2004. *Language and the Law*. Oxford: Blackwell.
- Goodrich, Peter. 1990. *Legal Discourse: Studies in Linguistics, Rhetoric and Legal Analysis*. London: Macmillan.
- Halliday, M.A.K. 1979. *Language As a Social Semiotic*. London: Arnold.
- Haugen, Einar. 1966 [1972]. "Dialect, Language, Nation". *Sociolinguistics: Selected Readings* ed. by John B. Pride & Janet Holmes, 97–111. Harmondsworth: Penguin.
- Hudson, Jean. 1998. *Perspectives on Fixedness: Applied and Theoretical*. Lund: Lund University Press.
- Jespersen, Otto. 1924 [1976]. "Living Grammar", in *The Philosophy of Grammar*. London: George Allen & Unwin, 17–29. Reprinted in *Readings in the Theory of Grammar* ed. by D.D. Bornstein, 82–93. Cambridge MA, Winthrop Publishers 1976.
- Kopaczyk, Joanna. 2009. "(Multi-word) Units of Meaning in 16th-century Legal Scots". *Selected Proceedings of the 2008 Symposium on New Approaches in English Historical Lexis (Hel-Lex 2)* ed. by Rod W. McConchie, Jukka Tyrkkö & Alpo Honkaphoja, 88–95. Somerville, MA: Cascadilla Proceedings Project.
- Kopaczyk, Joanna. 2011. "Standaryzacja tekstów w perspektywie historycznej. Analiza zbitek leksykalnych" [Text-type Standardisation in a Historical Perspective. Analysing Lexical Bundles]. *Metodologie językoznawstwa. Od ontologii do pragmatyki* [Linguistic Methodology. From Ontology to Pragmatics]. ed. by Piotr Stalmaszczyk, 155–174. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
- Kopaczyk, Joanna. Forthcoming a. "Formulaicity in Scots Historical Corpora and the Lexical Bundles Method". *New Methods in Historical Corpus Linguistics* (Corpus Linguistics and Interdisciplinary Perspectives on Language 2). ed. by Paul Bennett, Martin Durrell, Silke Scheible & Richard J. Whitt. Tübingen: Narr.
- Kopaczyk, Joanna. Forthcoming b. "Applications of the Lexical Bundles Method in Historical Corpus Research". *PALC Proceedings* ed. by Piotr Pęzik. Frankfurt am Main: Peter Lang.
- Kopaczyk, Joanna. Forthcoming c. *Standardising Legal Discourse: The Language of Scottish Burghs 1380–1560*. Oxford Studies in Language and the Law. Series ed. Roger W. Shuy. Oxford: University Press.
- Mellinkoff, David. 1963. *The Language of the Law*. Boston, Toronto: Little, Brown and Company.
- Moon, Rosamund. 1998a. *Fixed Expressions and Idioms in English*. Oxford: Clarendon Press.
- Moon, Rosamund. 1998b. "Frequencies of Forms of Phrasal Lexemes in English". *Phraseology* ed. by Anthony P. Cowie, 79–100. Oxford: Clarendon Press.
- Pryde, George S. 1965. *The Burghs of Scotland: A Critical List*. London, New York: Oxford University Press.
- Tiersma, Peter M. 1999. *Legal Language*. Chicago: University of Chicago Press.
- Wang, Shih-Ping. 2005. "Corpus-based Approaches and Discourse Analysis in Relation to Reduplication and Repetition". *Journal of Pragmatics* 37.505–540.
- Williams, Christopher. 2005. *Tradition and Change in Legal English: Verbal Constructions in Prescriptive Texts*. Bern: Peter Lang.
- Wray, Alison. 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Wright, Laura, ed. 2000. *The Development of Standard English: 1300–1800; Theories, Descriptions, Conflicts*. Cambridge: Cambridge University Press.

Appendix

Folder codes

Aberdeenshire	ABD	literary	LITX
Angus	AGS	Midlothian	MLO
Ayrshire	AYR	Nairn	NRN
Berwickshire	BWK	Orkney	ORK
Dumfriesshire	DMF	Peebles	PBL
Dunbar	DNB	Perthshire	PTH
Edinburgh	EDB	Renfrewshire	RNF
East Lothian	ELO	Stirlingshire	STG
Fife	FIF	West Lothian	WLO
Lanark	LNK	location unspecified	XLOC
Laws	LAWS		
