

TWIN LEXICAL COLLOCATIONS IN LEGAL LATE MIDDLE ENGLISH¹

LUIS IGLESIAS-RÁBADE

University of Santiago de Compostela

ABSTRACT

The aim of this study is an attempt to examine the occurrence and expansion of lexical clusters (collocations) in legal late Middle English and to determine whether or not these multi-word items are recurrent, cohesive, arbitrary and domain-dependent lexical clusters. The definition and categorisation of collocations are analysed on the basis of the statistical probability of co-occurrence of some vocabulary items (textual perspective). For this purpose non-technical English texts of late Middle English were chosen that provided us with a common corpus that operated as a point of reference. Likewise a smaller body of legal texts of the same period was collected. The *WordSmith* program was used to create *word lists* of the two corpora and compute the *key words* of the legal corpus. This study concentrates on twin lexical collocations (e.g. *geue & bequeath*) found in the first 100 salient words of the legal corpus as it is assumed here that a significantly frequent lexical word in the legal corpus plays an important role in collocational patterns.

1. Introduction

The access to large amounts of real authentic data in computational lexicography and the development of multi-word extraction techniques have made possible the growth of corpus-based studies in modern phraseology. Most researchers in this field attempt to solve the key problems which multi-word expressions present in the applications of modern intercultural communication, professional translation, natural language generation, computational lexicography, machine translation, etc. In contrast, the study of multi-word items in earlier periods of languages has attracted less attention, particularly because the range of applications is also more limited. However, it is essential to investigate the origin and

¹ This research has been funded by the Spanish Ministry of Education and Science, grant number HUM2005-00562/FILO. This grant is hereby gratefully acknowledged.

development of modern multi-word items in order to understand the process of their formation and function.

The specific purpose here is an attempt to scrutinise the occurrence and expansion of lexical clusters (collocations) in legal late Middle English and to determine whether or not these multi-word items are recurrent, cohesive, arbitrary and domain-dependent lexical clusters, as Smadja (1993: 143) claims in the case of modern English.

Collocations are included in the complex and imprecise area of multi-word expressions, which have been studied from various theoretical perspectives and with different purposes. The definition and categorisation of collocations are still controversial. Thus, emphasis may be given to (a) the statistical probability of co-occurrence of some vocabulary items (textual perspective); (b) the syntactic and semantic rules which govern the association of these items (Grossmann and Tutin 2003; Choueka 1988; Gitsaki 1996; Mel'cuk 1988); or (c) the pragmatic and rhetorical functions of the items which co-occur (Moon 1994; Nattinger and DeCarrico 1992: 36; Gledhill 2000).

The present study is formulated within the framework of the first approach by categorising collocations in terms of co-occurrence and recurrence. Thus, the degree of probability of two or more items co-occurring and the frequency with which these two or more items occur in a language provide the notion of collocation in a great number of corpus-based studies. Some of these are summarised as follows: "a recurrent co-occurrence of words" (Clear 1993: 277); "recurrent combination of words that co-occur more often than expected" (Smadja 1993: 143); "arbitrary and recurrent word combination" (Benson 1990: 23); "a sequence of words that occurs more than once in identical form" (Kjellmer 1987: 133); "the way individual words co-occur with others" (Lewis 1993: 93). This statistical approach is already perceived in Halliday's characterization of collocation as the "syntagmatic association of lexical items" (1961: 276). Firth's "lexical composition approach" (1951 [1957]:196) is illustrated by Gitsaki as follows: "words receive their meaning from the words they co-occur with" (1996: 10).

The hypotheses formulated in this paper are threefold: firstly, the lexis of legal late Middle English is assumed to be significantly different from that of common late Middle English; secondly, the lexis of legal late Middle English is expected to be collocationally primed (a) to co-occur with other vocabulary items (eg. *lawe day*), (b) to prefer a particular grammatical role or function (e.g. one element acting as head and the other/s as modifier); and (c) to convey a certain semantic involvement (eg. *foul* is associated with something "unpleasant") (Hoey 1997, 2000); and thirdly, collocational priming is assumed to develop and change in the course of time.

2. Method and criteria for twin lexical collocations selection

I have chosen a system of extracting and analysing data based on the following method:

- a. Significant frequency data are used as the basis of my analysis. The *WordSmith Tools 4* computer program was used to create *word lists* and compute the *key words* of the two corpora designed below. WordSmith statistically presents the figures of a smaller corpus and compares it with a larger, reference corpus. So the program shows words which are detected in the smaller corpus more or less frequently than what one may guess on the basis of the frequency provided for such words in the larger, reference corpus.
- b. From the possible list of compiled collocations I filter out the erroneous ones, following the method of Mutual Information (MI) proposed by Church and Hanks (1991) and provided by the WordSmith program. Mutual Information compares and equalizes the probability of two words occurring joined with the probability of these words occurring independently. If two items x and y have probabilities of occurrence $p(x)$ and $p(y)$, their mutual information $MI(x,y)$ is formulated as

$$MI(x,y) = \log_2 \frac{p(x,y)}{p(x) \cdot p(y)}$$

When $p(x,y) = p(x) \cdot p(y)$ and the resulting value of $MI(x,y)$ is 0, it evinces that the two items are not significantly recurrent to form a collocation. Whenever $MI(x,y)$ is < 0 , then we assume that the two terms (x, y) are mutually complementary and form a collocation.

- c. The WordSmith also provides a lemmatization process. Using this tool I have included under the same lemma the various spelling and morphological forms of a given word. For example, by checking the collocation *landes and tenementes* in the legal corpus the Concord program detected 117 hits of *land*/lond* and tenem**. The distribution of forms is as follows: on the one hand the program exhibits 69 occurrences of *landes and tenementes*, 4 tokens of *land and tenementes* and 1 token of *landys and tenementys*, and, on the other hand, 40 tokens of *londes and tenemente*, 2 tokens of *londys and tenementys* and 1 token of *lond and tenements*. In like manner, the allocation of this collocation in the 7 subgenres, which is basically dialectal, is shown in Table 1:

Table 1. *Landes/londes and tenementes* type

File	<i>Landes and tenementes</i>			<i>Landes and tenementes</i>	
	Words	Hits	%	Hits	%
Lincoln Doc	75.589	52	0.69	3	0.05
Indentures	4.768	2	0.42	9	1.89
Wills	40.370	8	0.20	15	0.37
Seals	28.491	2	0.07	1	0.4
Gyls	209.483	10	0.05	0	0
Petitions	47.630	0	0	14	0.29
Depositions	2.067	0	0	1	0.49
Total	408.398	74		43	

- d. Finally, for the sake of setting limits to this study I have confined my research to four categories of lexical *and*-concordances: (i) N and N (e.g. *landes and tenementes*), (ii) V and V (e.g. *haue and hold*), (iii) Adj and Adj (e.g. *god and lawful*) and (iv) Adv and Adv (e.g. *unduely and ungodly*). It is important to note that when the element after *and* includes a term different from the category of the first term, the one before *and*, the collocational type is left out. For example, the collocational patterns *the kyng and the queene* or *the kyng and the worchepeful lordes* are not included under the category of N and N in this study.

It is also worth noting that I have included an *and*-concordance as a collocation when it had also fulfilled all the following criteria (cf. Moon's 1994 recognition of Multi-Word Items):

- Institutionalisation*. I have examined the degree to which an *and*-concordance is conventionalized.
- Fixedness*. An appraisal of the degree to which an *and*-concordance might be frozen has been carried out for each type.
- Non-compositionality*. All *and*-concordance types were inspected so as to determine the degree to which they could be interpreted as having a specialised unitary meaning.
- Recurrence*. The inclusion of an *and*-concordance has been settled in terms of recurrence. Thus the cut-off of occurrence (minimum frequency) has been set out at least 4 occurrences.

One term of the collocation passes this limit by counting all the occurrences of its variant-forms (spelling variants or grammatical forms) as lemmatisation has

been carried out. For example, the record of the most frequent form of *forme and condicion* (5 occ.), includes the other less common forms of both terms of the collocation, such as *fforme and condicion* (2 occ.), *fourme and condicio* (1 occ.), *formes and condicions* (1 occ.), *forme and condicione* (1 occ.), *forme and condicions* (1 occ.) and *forme and condycion* (1 occ.). Thus the lemmatised collocation *forme and condicion* is registered as having 12 occurrences.

There is no restriction for the inclusion of a collocation with regard to the number of subgenres in which it appears. As will be shown in the section of analysis of data, 52 collocation types of the legal corpus out of 124 occur in a single subgenre.

3. Corpora and data analysis

Firstly I have chosen non-technical English texts of late Middle English that provided me with a *Common Corpus* (*CoCo*, henceforth) that operated as a point of reference. Secondly I have gathered a smaller body of legal texts of the same period (*Legal Corpus*, *LeCo* henceforth). *CoCo* and *LeCo* have been designed chiefly on the basis of two simple criteria: (a) a "medium-oriented choice": the texts were chosen on the basis of their electronic readability. For this purpose, the Humanities Text Initiative, a unit of the University of Michigan's *Digital Library Production Service*, has provided me online access to full text resources of the *Corpus of Middle English prose and verse*; and (b) a general "topic-oriented choice": *CoCo* texts have been selected on the basis of their common (non-technical) character. Thus the type of text that has been chosen might have represented the common speech of the fifteenth century: fiction, drama and religion texts. Table 2 shows the texts and the number of words of the reference corpus:

Table 2. *CoCo* texts

Texts of <i>CoCo</i>	Words
<i>The Canterbury Tales</i>	206734
<i>Everyman</i>	8118
<i>Confessio Amantis</i>	241707
<i>Orpheus and Eurydice</i>	27703
<i>The minor poems of Robert Henryson</i>	5392
<i>The morall fabillis of Esope the Phrygian</i>	5346
<i>The testament of Cresseid</i>	7016
<i>The vision of Piers Plowman</i>	79767

<i>Octavian</i>	11627
<i>Pearl</i>	8418
<i>Sir Gawain and the Green Knight</i>	21270
<i>The alliterative Morte Arthure</i>	42112
<i>The siege of Jerusalem</i>	12044
<i>The Towneley plays</i>	99943
<i>The York plays</i>	99736
<i>Troilus</i>	75014
<i>Rewle</i>	15736
<i>Merlin</i>	293332
Total	1,261.015

LeCo has been designed on the basis of its restricted legal technical character² and it has been also divided into seven subgenres: (a) Depositions, (b) Lincoln Documents, (c) Gilds, (d) Indentures, (e) Petitions, (f) Signet and Privy Seals, and (g) Wills.

Table 3 provides the data of *LeCo* including the number of words of each subgenre:

² *LeCo* texts include four subcorpora:

- English gilds: the original ordinances of more than one hundred early English gilds: together with The olde Usages of the cite of Wynchestre; the Ordinances of Worcester; the Office of the Mayor of Bristol; and the Costomary of the Manor of Tottenhall-Regis: from manuscripts of the fourteenth and fifteenth centuries* (London: Early English Text Society, Oxford University Press, 1870, reprinted 1963).
- An anthology of Chancery English*. Compiled by John H. Fisher, Malcolm Richardson and Jane L. Fisher (Knoxville: University of Tennessee Press, 1984).
- Lincoln diocese documents, 1450-1544* (London: Kegan Paul, Trench, Trübner & Co., Ltd. and Humphrey Milford, Oxford University Press, 1914).
- Fifty earliest English wills in the Court of Probate, London: A. D. 1387-1439*. Church of England. Province of Canterbury. Prerogative Court. (London, New York and Toronto: Oxford University Press, 1964).

Table 3. *LeCo* wordlists of subgenres

Section/subgenre	Words	%
Depositions	2.067	0.50
Documents	75.589	18.50
Gilds	209.483	51.29
Indentures	4.768	1.16
Petitions	47.630	11.66
Signet and Privy Seals	28.491	6.97
Wills	40.370	9.88
Total	408.398	100

Apart from these internal characteristics of the corpora, it is worth noting that *CoCo* and *LeCo* show different external contextual features. Thus *CoCo* includes wide-ranging linguistic functions (informative, instructional, persuasive, etc.), different styles or prototypical text categories (expository, narrative, imaginative, etc.), different non-technical settings (formal, informal, intimate equal/down/up, distant down/up, interactive, etc.), different types of text (drama, correspondence, fiction, history, romance, etc) and different topics (religion, fiction, etc.). However, *LeCo* external contextual characteristics are more restrictive. Thus the texts are functionally informative and show both a statutory style and a formal, professional and distant down setting. The type of text is also restricted to civil law and official documents and, with regard to topic, the texts show an amalgamation of legal texts (documents, parliament petitions, indentures, wills, etc.).

The Wordlist program (WordSmith) calculated 1,261,015 tokens (running words) for the total word reckoning of *CoCo*, whereas for *LeCo* the program counted 408,398 tokens. The wordlists include tokens and rates, but it is important to identify properly "types" of tokens/words. For the purpose of this work I have included under the same type both the orthographic and morphological forms of a word. The program has calculated 67155 types for *CoCo* and 26032 for *LeCo*.

4. *LeCo* and-lexical concordances: Types and tokens

All lexical *and*-concordances found in *LeCo* and attested as collocations are shown in the Appendix which also includes the figures for these collocational patterns in *CoCo*, so we may infer the acceptance of these technical collocations in other common registers. It is worth noting, however, that only those *and*-concordances authenticated at least 4 times are included in *LeCo* as it is the minimum frequency to consider a given type as a collocation in this research.

However, all instances (from 1 to *n*-number) are counted in *CoCo* because when an *and*-concordance has been verified as a collocation in the legal corpus, a single appearance in the other non-technical register permits us to believe that a specialised collocation has spread to other non technical domains.

It is important to note that my investigation concentrates exclusively on lexical *and*-concordances (lexical word + *and* + lexical word). For example, if we check the concordances of the word *kyng*, which appears in 53rd position in *LeCo* rank with 216 tokens and in the 15th place in the *LeCo*-salient word list, it exhibits several collocational types in *LeCo*: *þe Kyng and þe lords* (6 occ.), *þe kyng and his subgitts* (4 occ.), *þe kyng and þe queene* (4 occ.) or *þe kyng and his counsaile* (4 occ.). However, the collocational pattern of these types is not lexical word + *and* + lexical word.

The distribution of collocational categories in *LeCo*, including types and tokens, is shown in Table 4 (See also Appendix).

Table 4. Collocational categories

Collocational pattern	Number of types	Number of tokens
N and N	73	929
V and V	34	350
Adj and Adj	14	261
Adv and Adv	3	30
Total	124	1,570

As Table 4 shows, *LeCo* includes 124 collocational types which incorporate 1570 tokens. The most frequent collocational category is N and N with 73 types and 929 tokens. The most recurrent type is *landes/landes and tenements*. It is worth noting that *Landes/landes* appears as a salient word in 46th position in the corpus rank, however it turns up in first position when it is calculated as forming a collocation *Landes/landes & N* (cf Appendix). Thus its collocational distribution is as follows: *landes/landes and tenementes* (117 occ.), *landes/landes and rentes* (5 occ.), *landes/landes and possessions* (5 occ.), and *landes/landes and subgiettes* (4 occ.).

Under the category V and V the legal corpus exhibits 34 types which incorporate 350 tokens. *Couenauteth and graunteth* with 32 tokens is the most recurrent *and*-concordance in this class and it appears in 8th position in the collocation type rank. However, the second highest V and V collocation (*haue and hold*) is placed in 13th position and includes 23 tokens (cf. Appendix).

The collocational pattern Adj and Adj contains 14 types and 261 tokens. The most frequent type is *trusti and wellbeloued* which is found on 131 occasions in

LeCo. It is worth mentioning that although *trusty* appears as a salient word in *LeCo*, in 61st rank position, however it turns up in first position when it is calculated as forming the collocation *trusti and wellbeloued*. It is also noteworthy that all instances were registered within the sequence (*right*) *trusty and wellbeloued*.

With regard to the collocational pattern Adv and Adv, the type *well and truly* is the most frequent with 16 occurrences; only two other types, *well and sufficiently* (10 occ.) and *unduely and ungodly* (4 occ.), were attested in the corpus.

An initial result worth mentioning is that my survey of *and*-concordances in *LeCo* such as *fraternite & crafte*, *day & yere*, *make & ordeyne*, *geue & bequeth* etc. provides a list of 124 collocational types which include 1570 tokens. It is also interesting to note that the minimum frequency of a collocational type to occur has been set at least 4 occurrences.

5. Collocational patterns in *LeCo* salient lexical words

LeCo salient words are statistically significant items of higher frequency in *LeCo* than in *CoCo*. I assume here that a significantly frequent lexical word is expected to play some role in collocational patterns. Thus, the analytical procedure continues with the scrutiny of “key/salient words” in *LeCo* comparing them with their occurrence in *CoCo*. The Keyword program shows 392 salient items in *LeCo*, exhibited in descending order of saliency. Table 5 exhibits an extract from the list of salient words which includes only the lexical words found in the hundred most salient items in *LeCo*. It is important to note that words at the top of the key word list are more frequent than those at the bottom.

Table 5. Salient words of *LeCo*

<i>LeCo</i> Rank	Lexical words saliency	Lexical word	Freq. <i>LeCo</i>	%	Freq. <i>CoCo</i>	%	Keyness	<i>p</i> -score
3	1	SAID	3.656	2,78	549	0,04	15.541,5	0,000000
14	2	PAIE	565	0,43	18		2.721,8	0,000000
24	3	YERE	649	0,49	257	0,02	2.269,1	0,000000
25	4	CHIRCHE	546	0,41	163	0,01	2.044,6	0,000000
27	5	BEQUETHE	402	0,31	8		1.970,8	0,000000
30	6	ALDERMAN	356	0,27	3		1.779,6	0,000000
31	7	FORESAID	350	0,27	4		1.740,1	0,000000
32	8	SOULE	402	0,31	58		1.709,2	0,000000
33	9	ORDEYND	317	0,24	0		1.615,0	0,000000
36	10	MAKE	351	0,27	51		1.490,8	0,000000
43	11	EXECUTOR	281	0,21	7		1.366,8	0,000000

44	12	FRATERNITE	266	0,20	2		1.331,9	0,000000
49	13	PORTION	232	0,18	0		1.181,8	0,000000
51	14	COMEN	268	0,20	50		1.096,7	0,000000
53	15	KYNG	216	0,16	10		1.020,0	0,000000
54	16	RESON	223	0,17	17		1.016,0	0,000000
55	17	DAY	801	0,61	2.082	0,13	1.014,6	0,000000
60	18	PARTE	318	0,24	195	0,01	970,5	0,000000
62	19	GEUE	191	0,14	2		951,0	0,000000
63	20	LORD	203	0,15	13		937,9	0,000000
64	21	TYME	662	0,50	1.550	0,10	926,5	0,000000
65	22	CYTE	283	0,21	141		925,4	0,000000
68	23	PRESENT	263	0,20	117		889,6	0,000000
70	24	TOWNE	263	0,20	119		885,3	0,000000
75	25	TERME	252	0,19	113		850,5	0,000000
76	26	TENEMENT	176	0,13	7		838,2	0,000000
80	27	TESTAMENT	201	0,15	41		810,4	0,000000
81	28	SUCCESSOUR	159	0,12	0		809,9	0,000000
86	29	HEIRES	190	0,14	45		745,6	0,000000
89	30	WELBELOUED	146	0,11	2		722,8	0,000000
96	31	CLERC	174	0,13	45		671,2	0,000000
97	32	MANER	188	0,14	71		665,0	0,000000
98	33	BODY	222	0,17	144		663,7	0,000000
100	34	CHARGE	213	0,16	144		627,0	0,000000

Although I have examined all the *and*-concordances in *LeCo* that have at least four occurrences and have fulfilled the other three criteria detailed above (i. *institutionalisation*, ii. *fixedness*, iii. *non-compositionality*) for an *and*-concordance to be considered a collocational pattern, my investigation in this section is basically restricted to the collocations formed by the 34 key lexical items which appear in the list of the first 100 salient words in *LeCo* as shown in Table 5. *LeCo* figures (types and tokens) have also been compared with those found in *CoCo* which appear in brackets.

- i) *said* (*LeCo* rank 3; Salient lexical word 1, Keyness = 15.541,5; $p = 0,000000$)

Said is the first *LeCo* salient word, but it is also a very frequent word of *LeCo* as it occurs 3.656 times and takes the third position in *LeCo* rank after *the* and *of*. It is repeatedly used in expressions such as *the said William*, *the said abbot*, etc.

However, the program has not spotted any collocational pattern with *said*, as the only *and*-lexical concordance detected in *LeCo* is (1):

- 1) Also it is ordeyned qwan the messe is *seid and ended*, (*Gilds*)

Seid and ended of example (1) does not meet any of the four criteria used in this study to determine the collocational framework of an *and*-concordance.

- ii) *paie* (*LeCo* rank 14; Salient lexical word 2; keyness = 2.721,8; $p = 0,000000$)

The second highest salient word in *LeCo* is *paie*. It appears in 14th position in *LeCo* rank as, after being lemmatised, it has been detected on 565 occasions. The concord program has spotted the collocation pattern *content and paid*, which includes 13 examples. This collocational type has no occurrences in the common corpus (*CoCo* = 0). Example (2) illustrates this collocational type:

- 2) that þe seid sume of CCCC marc be fully *content and paid* to the seid Robert or his certeyn Attorney (*Indenture*)

- iii) *yere* (*LeCo* rank 24; Salient lexical word 3; keyness = 2.269,1; $p = 0,000000$)

Yere appears in 24th position in *LeCo* rank with 649 tokens, but it appears in third position in the *LeCo* salient word list. The concord program has found the collocational type *day and yere* 35 times. No instances of *yere* were found as left-collocate. One very frequently finds this collocational pattern to indicate the precise date on which a legal action has been performed, signed, etc. or to indicate that an action with legal implications is going to be carried out in a future time. It is also interesting to note that no occurrences of *day and yere* were detected in the common corpus (*CoCo* = 0). Let's consider example (3):

- 3) Yeven the *daye and yere* aboute Writen (*Lincoln Doc.*)

- iv) *chiriche* (*LeCo* rank 25; Salient lexical word 4; keyness = 2.044,6; $p = 0,000000$)

Chiriche follows in the *LeCo* salience scale in fourth position. It is also very common in legal texts as it appears in 25th position in *LeCo* rank and includes 546 occurrences. The reason why *chiriche* is much more frequent in *LeCo* than

in *CoCo* is that many legal documents had some relation with a given parish church that acted as landlord or tenant and therefore it is subject to many legal transactions. However, *chiriche* shows no *and*-lexical concordances, whether as left or right-collocate.

- v) *bequeth* (*LeCo* rank 27; Salient lexical word 5; keyness = 1.970,8; $p = 0,000000$)

Bequeth appears in fifth position in the saliency word list and occupies 27th place in *LeCo* rank with 402 tokens. It exhibits a single collocational type, *geue and bequeth*, which includes 17 tokens and it is restricted to legal domains as it has no occurrences in the common corpus (*CoCo* = 0). Let's consider example (4):

- 4) and the same howse I *geue and bequeyeth* to cuthbert my Sone (*Lincoln Doc.*)

- vi) *alderman* (*LeCo* rank 30; Salient lexical word 6; keyness = 1.779,6; $p = 0,000000$)

The 6th salient word, *alderman*, takes 30th position in *LeCo* rank with 356 tokens. Its use is restricted to the plural form and to the subgenre of *Gylde*. It is profusely employed in quasi-formulaic expressions addressing the members of a guild such as (*ye*) *aldermen and* (*ye*) *bretheren*. It occurs in 7 different types of collocational patterns. Thus it collocates with (*gyild*) *brothers* on 37 occasions (*CoCo* = 0), with *maistres* (*LeCo* 13 occ., *CoCo* 0), with *inhabitant*es (*LeCo* 6 occ., *CoCo* 0), with *bedel* (*LeCo* 6 occ., *CoCo* 0), with *felas* (*LeCo* 5 occ., *CoCo* 0), with *clerc* (*LeCo* 5 occ., *CoCo* 0), *skewe* (*LeCo* 5 occ., *CoCo* 0). *Alderman* always appears as left-collocate. It is interesting to note that no occurrences were found of any type in the common corpus. Let's consider (5):

- 5) and preyer shal bene reherside and seyde at euery tyme yat (*ye*) *alderman and* (*ye*) *bretheren* bene togedere (*Gilds*)

- vii) *forsaid* (*LeCo* rank 31; Salient lexical word 7; keyness = 1.740,1; $p = 0,000000$)

Forsaid continues as the 7th most *LeCo*-salient lexical word and in 31st position in *LeCo* rank with 350 tokens. It is widely used in all subgenres of the legal texts of the corpus (matching together with *abouesaid*) to refer to given information provided in a previous part of the document. The concord program does not exhibit any type of collocational patterns covered by this research, neither in *LeCo* nor in *CoCo*.

- viii) *soule* (*LeCo* rank 32; Salient lexical word 8; keyness = 1.709,2; $p = 0,000000$)

The next *LeCo*-salient lexical word is *soule* which rises to 32nd position in *LeCo* rank with 402 occurrences. It displays the collocational type *body and soule* in 5 instances. *Soule* is always a right-collocate, both in *LeCo* and *CoCo*, so the concord program has not detected any occurrence of **soule and body*. It is important to remark that this collocational framework might have been extensively used in everyday intercourse as it appears in 10 instances in the common corpus (*CoCo* = 10). Consider (6):

- 6) yat godd of his mercy saue hem and kepe hem, *body and soule*, and yeue heme grace here (*Gilds*)

- ix) *ordeyne* (*LeCo* rank 33; Salient lexical word 9; keyness = 1.615,0; $p = 0,000000$)

Ordeyne is the 9th most *LeCo* salient word and takes 33rd position in *LeCo* rank with 317 examples. This word had acquired a non-technical domain as it is frequently used in the common corpus. It exhibits two collocational types in *LeCo*: *ordeyne and make* (24 occ.) and *make and ordeyn* (23 occ.). So it appears with similar frequency as a right and left-collocate, but always with *make*. Although both terms of these collocational types were very common words, the concord program has not provided any instances of these collocational patterns in the common corpus (*CoCo* = 0 for both types). Let's consider (7) and (8):

- 7) for the performance of this my will and to pay my debtes, whome I *ordeyne and make* my Executrice and Mr. John Hasilwood (*Lincoln*)

- 8) Of this my Testament I *make and ordeyn* the said Alys my wyff my Executrice alone. (*Lincoln Doc.*)

- x) *make* (*LeCo* rank 36; Salient lexical word 10; keyness = 1.490,8; $p = 0,000000$)

The next most *LeCo*-salient lexical word (10th position) further down is *make* which is in 36th place with regard to *LeCo* rank, with 351 occurrences. What has been said for *ordeyne* above applies for *make*. Thus it shows two collocational types: *ordeyne and make* (24 occ.) and *make and ordeyne* (23 occ.) as in (7) and (8). Although *make* is a very common word it does not exhibit any instance of any of these two collocational patterns in common speech (*CoCo* = 0 for both types).

- xi) *executor* (*LeCo* rank 43; Salient lexical word 11; keyness = 1.366,8; $p = 0,000000$)

Executor comes next in the saliency rank scale (11th position) and it is exhibited in 43rd position in *LeCo* rank, with 281 occurrences. It shows three collocational types: *Executours and assignes* (23 occ.), *heiress and executours* (6 occ.) and *feoffes and executours* (4 occ.) as in (9), (10) and (11). *Executor* is restricted to legal domains as it has not been spotted in the common corpus, and therefore it has no examples of collocative patterns (*CoCo* = 0 for the thee types).

- 9) and with the said Thomas berrett prebendary that the said william his *executors and assignes* duryng all the said terme shall yerely content (*Lincoln*)
 10) to the said hugh Arthure and william their *heirs and executours* by thes endentures In witnesse wherof (*Indenture*)
 11) and that ther be a sufficient persone ordeyned by the aduysse of my said *feoffes and Executours* to receyve all the Issues and profutes of the said.. (*Lincoln Doc.*)

- xii) *fraternite* (*LeCo* rank 44; Salient lexical word 12; keyness = 1.331,9; $p = 0,000000$)

Fraternite appears in 12th place in the *LeCo*-salient word list and in the 44th position in *LeCo* rank with 266 tokens. Almost all these instances were used in the subgenre of *Gilds*, so it is properly a technical term but it might have been very common in the daily intercourse of guilds members. *Fraternite* shows two collocational types: *fraternite and gylde* (24 occ.) and *fraternite and crafte* (9 occ.) as in (12) and (13). No occurrences were found in the common corpus (*CoCo* = 0 in both types).

- 12) bretheren and susteren, and alle cristene, and for sustentacion of the *fraternite and Gylde* forseyd (*Gilds*)
 13) that the Master and Wardons of the *fraternitie and crafte* of Taylors and successors, shall use.. (*Gilds*)

- xiii) *porcion* (*LeCo* rank 49; Salient lexical word 13; keyness = 1.181,8; $p = 0,000000$)

The next *LeCo*-salient lexical word (13th position) is *porcion*. It occurs in 48th place in *LeCo* rank scale with 232 occurrences. Four instances were found forming the collocational type *porcion and parte* as shown in (14). *Porcion*

appears as left-collocate in all examples. No occurrences were exhibited in the common corpus for this type (*CoCo* = 0).

- 14) and his successours to entre and distreyne in all and euerye *porcion and parte* of the said.. (*Lincoln Doc.*)

- xiv) *come* (*LeCo* rank 51; Salient lexical word 14; keyness = 1.096,7; $p = 0,000000$)

Come appears as the 14th most *LeCo*-salient lexical word and in 51st position in *LeCo* rank with 268 tokens. *Come* is obviously an everyday word used in common conversation; however, by forming the collocational type *comen and offeren* as in (15) the pattern is restrictively used in technical domains. Thus the 11 occurrences of this collocational framework in *LeCo* had no counterpart in the common corpus (*CoCo* = 0).

- 15) and warne alle ye Gylde breyeren and sisteres to *comen and offre* an hal-peny at ye kyrke (*Gilds*)

- xv) *kyng* (*LeCo* rank 53; Salient lexical word 15; keyness = 1.020,0; $p = 0,000000$)

Kyng appears in 53rd position in *LeCo* rank with 216 tokens, but in 15th place in the *LeCo*-salient lexical word list. It is widely used in both *LeCo* and *CoCo* texts. It exhibits many collocational types (*þe Kyng and þe lords*, *þe kyng and his subgitts*, *þe kyng and þe queene* or *þe kyng and his counsaile*), although neither of them are covered by this research as both terms of the collocational types are preceded by a determiner as in (16):

- 16) þen þe said william shall aduertise *þe kyng and his counsaile* þerof and þey to pourueye him (*Indenture*)

- xvi) *reson* (*LeCo* rank 54; Salient lexical word 16; keyness = 1.016,0; $p = 0,000000$)

The next most *LeCo*-salient lexical word (16th position) is *reson* which appears in 54th place with regard to *LeCo* rank, with 233 occurrences. The program has not identified any *and*-lexical concordance for this word.

- xvii) *day* (*LeCo* rank 55; Salient lexical word 17; keyness = 1.014,6; $p = 0,000000$)

The following *LeCo*-salient lexical word in the legal corpus is *day* (17th salience position). As is shown in Table 5 it appears in 55th place in *LeCo* rank and has 801 occurrences. *Day* is a very common word and the two collocational types found in *LeCo* were unevenly represented in *CoCo*. Thus the concord program has found 35 occurrences of *day and yere* in *LeCo* (*CoCo* = 0). In like manner, *day and houre* occur on 5 occasions in *LeCo* and only on 1 in *CoCo*. *Day* is always a left-collocate in all examples, as in (17) and (18):

- 17) Yeven the *daye and yere* about Writen (*Lincoln*)
 18) they appere in the yeld halle, at the *day and houre* limited by the seid Bailies (*Gilds*)

xviii) *parte* (*LeCo* rank 60; Salient lexical word 18; keyness = 970,5; $p = 0,000000$)

The next *LeCo*-salient lexical word (18th position) further down is *parte*. It occurs in 60th place in *LeCo* rank scale with 318 occurrences. It exhibits two collocational types: *parte and parcel* and *porcion and parte*. Both types always maintain this structure, so in the 4 instances of *parte and porcion* in *LeCo*, *parte* is always left-collocate, whereas the four occurrences of *porcion and parte*, the latter is always right-collocate, as is shown in (19) and (20) respectively. These collocational patterns are basically used in technical legal domains as there are no instances in the common corpus (*CoCo* = 0).

- 19) to kepe vnto suche tyme as the saide yerely rente aud euery *parte and parcel* thereof with thearreragies of the same be fully (*Lincoln*)
 20) and his successours to entre and distreyne in all and euerye *porcion and parte* of the said ... (*Lincoln Doc.*)

xix) *geue* (*LeCo* rank 62; Salient lexical word 19; keyness = 951,0; $p = 0,000000$)

Geue continues as the 19th most *LeCo*-salient lexical word and in 62nd position in *LeCo* rank with 191 instances. It shows the collocational pattern *geue and bequeath*, which includes 17 tokens, as shown above in (4). It appears as left-collocate in every instance. It also seems to be restricted to technical fields as no instances were spotted in the common corpus (*CoCo* = 0).

xx) *lord* (*LeCo* rank 63; Salient lexical word 20; keyness = 937,9; $p = 0,000000$)

Following further down in the saliency rank scale is *lord* (20th place) which occurs in 203 examples, which gives it 63rd place in *LeCo* rank. *Lord* forms various collocational types such as *the kyng and the worchepeful lordes* with 6 occurrences in *LeCo* as in (21). However, the concord program has not detected any collocational type covered by this research.

- 21) he billes myne aduersaries replieden by mouthe and enformeden *the kyng and the worchepeful lordes* spirituelx and temperelx (*Chancery*)

xxi) *tyme* (*LeCo* rank 64; Salient lexical word 21; keyness = 926,5; $p = 0,000000$)

The next *LeCo*-salient lexical word is *tyme* which appears in 64th position in *LeCo* rank with 662 occurrences, whereas in *LeCo*-saliency scale it gets 21st position. It displays the collocational types *fourme and tyme* (4 occ.) and *hour and tyme* (4 occ.). In both types *tyme* is always a right-collocate, as in (22) and (23). Although *tyme* is a very frequent word in *CoCo* as well, it does not occur in any of these collocational patterns (*CoCo* = 0), basically because the context of these collocations is a formal document in which form and time are important terms with legal significance.

- 22) in suche *fourme and tyme* as pleseth to (*Chancery*)
 23) be *oure and time* assigned and ordeyned be the aldirman (*Gilds*)

xxii) *cyte* (*LeCo* rank 65; Salient lexical word 22; keyness = 925,4; $p = 0,000000$)

Cyte comes next in the saliency rank scale (22nd position) and it appears in 65th place in *LeCo* rank with 283 occurrences. It shows the collocational type *cyte and suburbes* on 7 occasions as in (24). It always appears as a left-collocate and is restricted to technical contexts as no occurrences were spotted in the common corpus (*CoCo* = 0):

- 24) alle the Londes and (tenements tat he hadde) in the *Cite and Suburbes* of London (*Chancery*)

xxiii) *present* (*LeCo* rank 68; Salient lexical word 23; keyness = 889,6; $p = 0,000000$)

The next *LeCo*-salient lexical word, *present* (23rd position) has 263 occurrences which gives it 68th place in *LeCo* rank. It does not show any *and*-lexical concor-

dances, neither in *LeCo* nor in *CoCo*.

xxiv) *towne* (*LeCo* rank 70; Salient lexical word 24; keyness = 885,3; $p = 0,000000$)

LeCo-saliency rank continues with *towne* in 24th position. It includes 263 occurrences which provides it with 70th place in *LeCo* rank. It shows a productive collocational type, *towne and marches*, with 12 instances in *LeCo*. Although it is also a very common word in *CoCo*, it does not exhibit any example of the collocational type described above (*CoCo* = 0). It is also noticeable that *town* is always a left-collocate as shown in (25):

25) all youre landes Rentes tenements with all oþer commodites of youre *towne and marches* afoersaide to þe moste prouffit in confermacon (*Seals*)

xxv) *terme* (*LeCo* rank 75; Salient lexical word 25; keyness = 850,5; $p = 0,000000$)

LeCo-saliency continues with *terme* which gets 25th position as salient word. *LeCo* includes 252 tokens raising its position in *LeCo* rank to 75th place. It exhibits the collocational type *ende and terme* on 8 occasions, as in (26). There are no examples in the common corpus (*CoCo* = 0) so this collocational pattern is restricted to legal domains. It is also worth noting that *terme* is always a right-collocate.

26) Cristemas next commyng aftir the date of these presentes vnto the *ende and terme* of twenty yeres (*Indenture*)

xxvi) *tenement* (*LeCo* rank 76; Salient lexical word 26; keyness = 838,2; $p = 0,000000$)

Tenement is the 26th most *LeCo*-salient lexical word and takes 76th position in *LeCo* rank with 176 examples, however it appears in second position when it is calculated as forming the collocation *landes and tenementes* which includes 117 occurrences as in (27). This collocative type is widely extended to all legal subgenres as it appears in all of them, but no occurrences are found in the common corpus (*CoCo* = 0). *Landes* is always a left-collocate.

27) Symon digby and his heires mailles peaseably to inioye the said *landes and tenementes* to them before appointed (*Lincoln Doc.*)

xxvii) *testament* (*LeCo* rank 80; Salient lexical word 27; keyness = 810,4; $p = 0,000000$)

LeCo-saliency rank continues with *testament* in 24th position. It includes 201 occurrences which gives it 80th place in *LeCo* rank. It exhibits a productive collocational type, *testament and will*, which appears on 20 occasions as in (28). There are no instances of this type in the common corpus (*CoCo* = 0) and the 20 examples detected in *LeCo* show that *testament* is always left-collocate. The type *will and testament* has been also attested in *LeCo*, but not in the number of occurrences (at least 4) that are required in this study to consider a given collocational pattern as a type.

28) and gyf myn Executours grace to make good ende of my *testament and wille*, and my feffe? also of my feffementes (*Wills*)

xxiii) *successour* (*LeCo* rank 81; Salient lexical word 28; keyness = 809,9; $p = 0,000000$)

Successour follows in *LeCo*-saliency rank in 28th position, and in the 81st in *LeCo* rank with 159 instances. It does not show any collocational type.

xxix) *heires* (*LeCo* rank 86; Salient lexical word 29; keyness = 745,6; $p = 0,000000$)

LeCo-saliency continues with *heires* (29th position) which rises to 86th place in *LeCo* rank with 190 occurrences. The concord program has not detected collocational types with this *LeCo*-salient word.

xxx) *welbeloued* (*LeCo* rank 89; Salient lexical word 30; keyness = 722,8; $p = 0,000000$)

Welbeloued appears as the 30th most salient lexical word and in 89th position in *LeCo* rank with 146 tokens. It includes the most frequent collocational type in *LeCo*, *trusty and welbeloued*, with 131 examples as in (29). No instances are exhibited in the common corpus (*CoCo* = 0). *Welbeloued* is always a right-collocate.

29) Be the kyng Worshipful fader in god our *trusty and welbeloued* / ffor certain causes of science. (*Chancery*)

xxxi) *clerc* (*LeCo* rank 96; Salient lexical word 31; keyness = 671,2; $p = 0,000000$)

Clerc follows in *LeCo*-saliency rank in 31st position and in the 96th in *LeCo* rank with 174 instances. It does not show any collocational type.

xxxii) *maner* (*LeCo* rank 97; Salient lexical word 32; keyness = 665,0; $p = 0,000000$)

LeCo-saliency continues with *maner* (32nd position) which rises to 97th place in *LeCo* rank with 188 occurrences. The concord program has not detected collocational types with this *LeCo*-salient word.

xxxiii) *body* (*LeCo* rank 98; Salient lexical word 33; keyness = 663,7; $p = 0,000000$)

Body follows in *LeCo* saliency scale in 33rd position. It appears in 98th position in *LeCo* rank and includes 222 tokens. It follows in *LeCo* rank scale and collocates with *soule* on 5 occasions as in (6). *Body* is always a left-collocate both in *LeCo* and *CoCo*. As has been said above, this collocational pattern includes 10 instances in the common corpus (*CoCo* = 10).

xxxiv) *charge* (*LeCo* rank 100; Salient lexical word 34; keyness = 627,0; $p = 0,000000$)

Finally *charge* is the 34th most *LeCo*-lexical salient word and appears at the end of the list of the top salient words (100th position) in *LeCo* rank with 213 tokens. It shows two productive collocational types: *costes and charges* with 30 occurrences and *charges and expenses* with 4. No examples are detected in the common corpus (*CoCo* = 0). It is important to note that *charges* is both a right and a left-collocate, as shown in (30) and (31) respectively:

30) to hold and occupye to suche tyme as the sayd Rent with *costes and chardgies* made aboute he same then be fully contentyd.. (*Lincoln*)

31) to the chauncell of the saide rector fealowes and scolers with the *chardgies and expenses* of synginge bredde, wyne, waxe.. (*Lincoln*)

Some conclusions are worth offering here. Out of the 124 collocational types (1570 tokens) found in the whole legal corpus, 48 types (40%) and 741 tokens (47,9%) are detected with the 34 most salient lexical words found in the first 100 *LeCo*-salient words. At the beginning of this section I had assumed that a significantly frequent lexical word would play an important role in a collocational framework. The data shown in Table 5 and commented in this section now clearly confirm my initial assumption, basically because some of these 34 *LeCo*-lexical salient words

do not occur in the common corpus. *Alderman*, for example, is one of the terms of 73 collocations in *LeCo*, but that word does not occur in *CoCo*.

6. Collocational patterns in legal subgenres

The examination of collocational types and tokens in the 7 subgenres (see Table 6) analysed in this research reveals an uneven distribution. Thus 52 (43.3%) out of 124 collocational types appear in a single subcorpus with 639 tokens (41.35% of all tokens). For example, *trusti and wellbeloued*, which is the most recurrent *LeCo* collocational type with 131 tokens, is detected only in the *Seal* subcorpus (cf. Appendix). It is also important to note that 39 types (32.5%) are found only in 2 subgenres and exhibit 375 tokens (24.27%). Thus, for example, *couenauteth and graunteth* occurs only in *Gilds* (9 occ.) and *Lincoln Doc.* (24 occ.) subgenres. More extensively used are the 17 types (14.1%) which occur in 3 subgenres with 163 instances (10.55%) such as *costes and charges*. However, only 7 types (5.83%) such as *issues and profites* are distributed in 4 subgenres and count for 118 occurrences (7.63%). Similarly, just 3 patterns (2.5%) such as *maner and fourme* are detectable in 5 subgenres, reckoning 98 tokens (6.34%). Only one type (0.83%), *day and yere*, is located in 6 subgenres and includes 35 occurrences (2.26%), and finally also a single pattern (0.83%), *landes and tenements*, is detected in all subgenres and it is very recurrent as it accounts for 117 collocations (7.57%).

Table 6. Collocational types in legal subgenres

N° of types	N° of tokens	% occ.	Subgenres
55	664	41.35	1
39	375	24.27	2
18	163	10.55	3
7	118	7.63	4
3	98	6.34	5
1	35	2.26	6
1	117	7.57	7
Total	124	1570	100

I assume the hypothesis that a collocational type is widely accepted in legal technical domains when it appears at least in 3 different subgenres. If so, only 29 collocational types, which occur in three or more subgenres, were actually productive at least in legal fields. The other 91 types out of 124 (65.62% of collocations), which appear in only one or two genres, had a restricted use even in the legal sphere.

7. The incidence of *LeCo* collocational types in *CoCo*

The data collection reveals that the greater part of *LeCo* collocational patterns do not occur in *CoCo* (cf. Appendix). In fact, only 12 *LeCo* types (8.53%) out of 124 were also detectable by the concord program in *CoCo*. Besides, the level of recurrence in *CoCo* is not statistically significant as these 12 *LeCo* types which had some occurrences in *CoCo* reckon only 30 tokens (1.32%). Furthermore, the number of *CoCo*-occurrences for 9 out of 12 *LeCo* types found in *CoCo* does not reach the level of recurrence which has been set (at least 4 tokens) in this study so as to consider it a collocational pattern. Thus only *body and soule* includes 10 tokens in *CoCo* in different texts, whereas *pray and require* (5), and all of them in a single text (*Merlin*), passes the cut-off of occurrence as Table 7 reveals:

Table 7. Types which appear in both *LeCo* and *CoCo*

<i>LeCo</i> Rank	Collocational type	<i>LeCo</i> tokens	<i>CoCo</i> tokens	<i>CoCo</i> -Texts
21	Well and trully	16	1	1
50	Gode and trewe	8	3	3
56	Ferme and stable	7	2	1
60	Pray and require	7	5	1
78	Day and place	6	1	1
90	Kepe and meyntene	5	1	1
95	Power and autorite	5	1	1
98	Saue and kepe	5	1	1
100	body and soule	5	10	7
104	Feith and conscience	4	1	1
112	Mercy and grace	4	3	3
Total		72	30	21

Assuming a similar rationale used for *LeCo* subgenres, I take for granted the hypothesis that a collocational type is widely accepted in common domains when it appears at least in 3 different texts. As Table 7 indicates only three *LeCo* collocational types, found in *CoCo*, *body and soule*, *gode and trewe* and *mercy and grace*, appear, each of them, in more than two texts. This, therefore, indicates that these three collocational types might be used both in colloquial and in technical (legal) domains. The other *LeCo* types found in *CoCo*

do not have sufficient recurrence to be considered collocational patterns of common use.

8. Conclusions

The attempt to show that legal late Middle English used some specific recurrent lexical clusters has been completely attested in this study. I have also proved that the meanings of these clusters present some characteristics of collocational patterns (non-compositionality, institutionalization and fixedness). The scope of this research is confined to the study of twin collocations joined by *and*; many collocational patterns which appear in my legal corpus will need separate discussion.

By comparing the wordlist of the legal corpus that I have gathered with the wordlist of a larger common corpus, I have found that 392 words were statistically significant items of higher frequency in the technical corpus. I had assumed that these salient words of the legal corpus were expected to play some role in collocational patterns. This assumption was completely witnessed. Thus, out of the 124 collocational types (1570 tokens) found in the whole legal corpus, 48 types (40%) and 741 tokens (47,9%) are detected only in the 34 lexical salient words found in the first 100 *LeCo*-salient words. I had also assumed the principle that a collocational pattern is extensively accepted in the legal sphere when it appears at least in 3 different subcorpora. My study shows that only 29 collocational patterns were exhibited in three or more subgenres, while the other 91 types out of 124 (65.62% of collocations) appear in only one or two subcorpora. This proves that few collocational types were actually productive within the legal domains.

What seems much more interesting is the fact that the greater part of the collocational patterns found in the legal corpus do not occur in the common corpus. Only 10 types (8.33%) out of 124 appear in the common corpus and only one type reaches the level of recurrence which could be considered statistically significant. Furthermore, only three collocational types, *body and soule*, *gode and trewe* and *mercy and grace*, detected also in the common corpus, appear in three or more different texts with sufficient frequency so as to be considered fully productive.

REFERENCES

- Baker, Mona – Gill Francis – Elena Tognini-Bonelli (eds.)
1993 *Text and technology*. Amsterdam: John Benjamins.
Benson, Morton
1990 "Collocations and general-purpose dictionaries", *International Journal of Lexicography* 3/1: 23-34.

- Choueka, Yakov
1988 "Looking for needles in a haystack", *Proceedings of Recherche d'Information Assistée par Ordinateur* 1988: 609-623.
- Church, Kenneth – Patrick Hanks
1989 "Word association norms, Mutual Information and Lexicography", *Computational Linguistics* 16/1: 22-29.
- Clear, Jeremy
1993 "From Firth principles. Collocational tools for the study of collocation", in: Mona Baker – Gill Francis – Elena Tognini-Bonelli (eds.), 271-292.
- Coulthard, Malcolm (ed.)
1994 *Advances in written text analysis*. London: Routledge.
- Firth, John R.
1968 "A synopsis of linguistic theory 1930-55", in: Frank R. Palmer (ed.), 168-205.
- Gitsaki, Christina
1996 *The development of ESL collocational knowledge*. [Unpublished Ph. D. dissertation, The University of Queensland, Brisbane, Australia].
- Gledhill, Chris J.
2000 *Collocations in science writing*. Tübingen: Gunter Narr Verlag.
- Grossmann, Francis – Agnès Tutin (eds.)
2003 *Les collocations: Analyse et traitement*. Amsterdam: De Werelt.
- Halliday, Michael A. K.
1961 *Categories of the Theory of Grammar*. Edinburgh: Edinburgh University Press.
- Hoey, Michael
1997 "From concordance to text structure: New uses for computer corpora", in: James Melia – Barbara Lewandowska (eds.), 2-23.
- Kjellmer, Göran
1987 "Aspects of English collocations", in: Willem Meijs (ed.), 133-140.
- Meijs, Willem (ed.)
1987 *Corpus linguistics and beyond*. Amsterdam: Rodopi.
- Melia, James – Barbara Lewandowska (eds.)
1997 *Proceedings of Practical Applications in Language Corpora*. Łódź: Łódź University Press.
- Mel'čuk, Igor
1988 "Semantic description of lexical units in an explanatory combinatorial dictionary: Basic principles and heuristic criteria", *International Journal of Lexicography* 1/3: 165-188.
- Moon, Rosamund
1994 "The analysis of fixed expressions in text", in: Malcolm Coulthard (ed.), 117-135.
- Nattinger, James R. – Jeanette S. DeCarrico
1992 *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Pitch, Herbert (ed.)
1996 *Orality and literacy in Early Middle English*. Tübingen: Gunter Narr Verlag.
- Palmer, Frank R. (ed.)
1968 *Selected papers of J. R. Firth 1952-1959*. London: Longman.
- Sinclair John M.
1991 *Corpus concordance collocation*. Oxford: Oxford University Press.

- Smadja, Frank
1993 "Retrieving collocations from text: Xtract", *Computational Linguistics* 19/1: 143-177.
- Schaefer, Ursula
1996 "Twin collocations in the Early Middle English *Lives of the Katherine Group*", in: Herbert Pitch (ed.), 179-98.

INTERNET SOURCES

- Evert, Stefan
2004 *The statistics of word co-occurrences: Word pairs and collocations*. [Ph. D. dissertations, IMS, University of Stuttgart], available at <http://www.collocations.de/EK/> (date of access: 2004).
- Nelson, Mike
2000 *A corpus-based study of Business English and Business English teaching materials*. [Unpublished Ph. D. dissertation, University of Manchester], available at <http://users.utu.fi/micnel/thesis/> (date of access: 2000).

APPENDIX

Rank	Collocation	Type	Subgenres of <i>LeCo</i>								Total <i>LeCo</i>	Total <i>CoCo</i>
			Ind	Pet	Wil	Seal	Dep	Gyl	Doc			
1	Trusti & well-beloued	Adj + Adj	0	0	0	131	0	0	0	131	0	
2	Landes & tenements	N + N	1	15	23	3	1	12	52	117	0	
3	Maister & wardens	N + N	0	0	0	0	0	80	0	80	0	
4	Baillies & communes/alty	N + N	0	0	0	0	0	60	0	60	0	
5	Maner & fourme	N + N	0	10	9	2	0	22	17	60	0	
6	Spiritual & temporel	Adj + Adj	0	31	0	3	0	3	2	41	0	
7	Day & yere	N + N	2	3	6	0	1	3	19	35	0	
8	Couenauteth & graunteth	V + V	0	0	0	0	0	9	24	32	0	
9	Costes & charges	N + N	0	6	0	0	0	7	17	30	0	
10	Testament & will	N + N	0	0	24	0	0	0	4	28	0	
11	Fraternite & gylde	N + N	0	0	0	0	0	24	0	24	0	
12	Ordeyne & make	V + V	0	2	3	1	0	3	15	24	0	
13	Maire & bailiffs	N + N	0	7	0	3	0	13	0	23	0	
14	Make & ordeyne	V + V	0	1	10	1	0	8	3	23	0	
15	Executors & assignes	N + N	0	0	1	00	0	0	22	23	0	
16	Haue & hold	V + V	0	0	0	0	0	0	23	23	0	
17	Dirige & masse	V + V	0	0	0	0	0	8	12	22	0	
18	Issues & profites	N + N	0	3	6	0	0	1	11	21	0	
19	Will & testament	N + N	0	0	2	0	0	0	17	19	0	
20	Geue & be-queth	V + V	0	0	1	0	0	1	15	17	0	

21	Well & trully	Adv+ Adv	3	1	0	0	0	0	9	3	16	1
22	Maister & brethern	N + N	0	0	0	0	0	0	16	0	16	0
23	Godes & catalles	N + N	0	3	4	1	0	0	5	2	15	0
24	Heiress & assignes	N + N	7	0	1	1	0	0	0	6	15	0
25	Rentes & profits	N + N	0	0	2	0	0	0	12	0	14	0
26	God & lawful	Adj + Adj	0	0	0	0	0	7	7	14	0	
27	Content & paid	Adj + Adj	1	0	0	0	0	0	0	12	13	0
28	Alderman & maisters	N + N	0	0	0	0	0	0	13	0	13	0
29	Forme & condicions	N + N	0	0	0	0	0	0	0	12	12	0
30	Liberate & allocate	V + V	0	0	0	12	0	0	0	0	12	0
31	Maire & sheriff	N + N	0	0	0	0	0	0	12	0	12	0
32	Town & marches	N + N	0	0	0	12	0	0	0	0	12	0
33	Yeld & pay	V + V	0	0	0	0	0	0	0	12	12	0
34	Wise & discret	Adj + Adj	0	10	0	1	0	0	0	0	11	0
35	Comen & offeren	V + V	0	0	0	0	0	0	9	2	11	0
36	Ordeyne & establishe	V + V	0	6	0	0	0	0	4	2	11	0
37	Felowes & scolers	N + N	0	0	0	0	0	0	0	10	10	0
38	Placebo & dririge	V + V	0	0	1	0	0	0	8	1	10	0
39	Prior & chanons	N + N	0	8	0	0	0	0	2	0	10	0
40	Well & sufficently	Adv+ Adv	0	0	0	0	0	0	1	8	9	0
41	Entre & distraîne	V + V	0	1	0	0	0	0	0	8	9	0
42	Fraternite & crafte	N + N	0	0	0	0	0	0	9	0	9	0
43	Liberte &	N +	0	5	0	3	0	1	0	0	9	0

89	Grete & notable	Adj + Adj	0	5	0	0	0	0	0	5	0
90	Helping & assisting	V + V	0	0	0	0	0	0	5	5	0
91	Kepe & meyntene	V + V	0	0	0	0	0	4	1	5	1
92	Maister & frends	N + N	0	0	0	0	0	5	0	5	0
93	Noble & gracious	Adj + Adj	0	5	0	0	0	0	0	5	0
94	Ordinaunces & ruelles	N + N	0	0	0	0	0	5	0	5	0
95	Personage & prebend	N + N	0	0	0	0	0	0	5	5	0
96	Power & autorite	N + N	0	1	0	0	0	3	1	5	1
97	Prebend & personage	N + N	0	0	0	0	0	0	5	5	0
98	Reuenues & profits	N + N	1	0	2	0	0	2	0	5	0
99	Saue & kepe	V + V	0	0	2	0	0	2	1	5	1
100	Body & soule	N + N	0	1	1	0	0	1	1	5	10
101	Day & heure	N + N	0	0	1	0	0	2	2	5	1
102	Porcion & parte	N + N	0	1	1	0	0	1	2	4	0
103	Fourme & tyme	N + N	0	0	0	0	0	2	2	4	0
104	Assigned & ordeined	V + V	0	1	0	0	0	3	0	4	0
105	Castel & lorship	N + N	0	4	0	0	0	0	0	4	0
106	Charges & expenses	N + N	0	1	0	0	0	0	3	4	0
107	Curate & kirke-warde	N + N	0	0	0	0	0	4	0	4	0
108	Feith & concience	N + N	0	4	0	0	0	0	0	4	1
109	Foeffes & executours	N + N	0	0	1	0	0	0	3	4	0
110	Forme & manner	N + N	0	1	0	0	0	2	1	4	0
111	Godes & ornaments	N + N	0	0	0	0	0	4	0	4	0

112	Heresies & errours	N + N	0	0	0	0	0	0	1	3	4	0
113	Lordshipe & touneshipe	N + N	0	4	0	0	0	0	0	0	4	0
114	Maner & touneshipe	N + N	0	4	0	0	0	0	0	0	4	0
115	Maner & condicioun	N + N	0	0	4	0	0	0	0	0	4	0
116	Mercy & grace	N + N	0	0	4	0	0	0	0	0	4	3
117	Power & diligence	N + N	0	0	0	0	0	0	4	0	4	0
118	Reparacioun & mantenaunc	N + N	0	0	0	0	0	0	0	4	4	0
119	Restore & deliuer	V + V	0	4	0	0	0	0	0	0	4	0
120	Singe & rede	V + V	0	0	4	0	0	0	0	0	4	0
121	Unduely & ungoodly	Adv+ Adv	0	4	0	0	0	0	0	0	4	0
122	Usage & custome	N + N	0	1	0	1	0	2	0	0	4	0
123	Use & proffit	V + V	0	0	0	0	0	0	2	2	4	0
124	Worthy & gracious	Adj + Adj	0	4	0	0	0	0	0	0	4	0
Total tokens			16	218	133	203	2	499	476	1570	30	