SOME LEXICAL COLLOCATIONAL PATTERNS IN LATE MIDDLE ENGLISH LEGAL TEXTS[1]

LUIS IGLESIAS-RÁBADE
University of Santiago de Compostela
iarabade@usc.es

Introduction

The recurrent use in modern languages of lexical and grammatical clusters whose meaning is not to be interpreted on a word-by-word basis, has recently attracted the attention of many scholars. However, the study of these clusters (collocations) in the earlier periods of languages has inspired less interest as there are no clear applications in modern intercultural communication (natural language generation, machine translation, etc.). It is worth noting, however, that all these clusters were formed at a given moment of a language's history and developed for a specific function. The purpose of this study is to attempt to scrutinize the influence and force of collocations in legal late Middle English, and to show how they either specialized in this technical area or slipped into common speech.

The reasons why some words tend to associate with others in a given order and at a given time in the history of a language (e.g. *trusty and welbeloued, goods and chattels*) have not been satisfactorily explained. Whatever the state of mental organisation turns out to be for speakers to associate certain words, it must have eventually developed in a cultural framework. Thus most of the collocations are dependent on culture and domain (usually technical). Therefore they should not be simply analysed as arbitrary recurrent items characterised only by a statistical probability of co-occurrence. Furthermore, from a grammatical perspective it seems

necessary to redefine the syntactic and semantic rules which govern these associations of words. Certainly, some theories and approaches have been proposed to explain the factors and rationale which determine the use of lexical and grammatical clusters. Thus, the notion of collocation has been studied from a pragmatic perspective, emphasising the rhetorical function of the multi-word expression in discourse. This level of study is particularly interesting because the analyst is involved in the stylistic contrast between marked forms such as "take notice" and the unmarked "notice". The difference is assessed by analysing rhetorical factors, as Moon (1994: 117) states: "fixed expressions represent meaningful choices on the part of the speaker/writer". Nattinger and DeCarrico (1992:36) differentiate unmarked collocations (co-ocurrences of lexical items) from lexical phrases or marked collocations (polywords, institutionalised phrases, phrasal constraints) as different choices of expression. Whereas collocations have a pragmatic function, the unmarked co-ocurrence of lexical items are expressions "that have not been assigned particular pragmatic functions by pragmatic competence". However, Gledhill (2000:16) maintains that a "normal text rarely moves in a clear-cut way from unmarked to marked expression [...] It is more realistic to picture a text as a sequence of different types of discourse signal".

Other scholars focus on the syntactic and semantic rules which govern the association of words. Thus, Grossmann and Tutin (2003) have examined collocations as pre-constructed syntactic units, Choueka (1988) has studied them as lexically determined elements of grammatical structures, and Gitsaki (1996:17) has emphasized the idea that word associations occur in patterns.

A different approach is proposed by those who analyse collocations from a lexical/textual point of view, emphasising the statistical probability of some co-occurring items. Following this method, Clear has studied collocations as a "recurrent co-occurence of words" (1993: 277); Smadja, as a "recurrent combination of words that co-occur more often than expected" (1993: 143); Benson, as an "arbitrary and recurrent word combination" (1990: 23); Kjellmer, as "a sequence of words that occurs more than once in identical form" (1987: 133).

My analysis of the collocational framework in late Middle English legal texts tries to follow a lexical description based on the analysis of 'collocation' and 'set' as counterparts of 'structure' and 'system' in grammatical analysis, emphasizing the collocational structure rather than the rules that operate within the set. Thus, following Carter (1987: 50), I shall select items from lexical sets instead of choosing types of grammatical structures. For the purpose of this work, 'collocation' is understood as a "probable co-occurrence of items" (Malmkjaer, 1991: 302), whereas 'set', which envisages instances of one and the same syntagmatic relation, is not analysed. Thus I shall not consider in this study the

syntagmatic relation which, for example, Halliday, (1966: 151-157) proposes when he considers "a strong argument", "he argued strongly", "the strength of his argument" and "his argument was strengthened" as syntagmatically related units. In Halliday's view "strong", "strongly", "strength" and "strengthened" collocate with "argue/argument". My data analysis is much simpler as it is basically concentrated on combinations of words which 'arbitrarily' associate with each other more often than expected, and where such association is presumed to be domain (legal) dependent.

Another aspect that is beyond the scope of this study is determining whether the lexeme has its own independent meaning or whether such a meaning is only shaped by the set of its collocations, as Nattinger & DeCarrico (1992: 181-182) have stated. Sinclair (1991: 115-116) also maintains that the "relative frequency of node and collocate determines whether the collocational relation will contribute to the meaning of the node." This seems to be clear in the case of composite predicates in which the support verb has been emptied of lexical meaning as the latter has been displaced to the deverbal noun (e.g. take notice, take care, take advantage, etc.). In broad terms, I have assumed in this study that the greater capacity a lexeme has to develop collocational patterns, the greater restrictions the node has.

Method and criteria for collocational patterns selection

The method used in this study to select collocational types is as follows:

a) I have employed the *WordSmith Tools 4* programme to draft *word lists* of two corpora (the specific/legal corpus and the common/reference corpus). WordSmith also provides a catalogue of key/salient words by contrasting statistically the number of tokens in both corpora. Thus the programme allows us to view a record of words which are identified in the specific legal corpus more frequently than in the common corpus.

b) Collocations were retrieved and filtered out applying Church and Hanks's Mutual Information (MI) technique which can be conducted by the Wordmith programme. Mutual Information contrasts and balances the probability of two words occurring mutually joined with the probability of these words occurring independently.[2]

c) The word lists which have been planned for the two corpora were sorted out through a process of lemmatization supplied by Wordsmith. Thus I have accumulated under the same lemma the spelling variants and morphological forms of the same word.

d) I have restricted my research to six categories of lexical collocations: i) $N^1 + N^2$

(e.g. *market day*), ii) N *and* N (e.g. *wil and testament)*, iii) Adj + N (e.g. *gracious lord*), iv) Adj *and* Adj + N (e.g. *trusti and wellbeloued lord*), v), V + N (e.g. *take profit*) and vi) V *and* V (e.g. *execute and perform*).

e) The categorization of collocations in this study is based on the following determining factors: a) the degree of probability of a multi-word item is measured in relation with its degree of *institutionalization* (conventionalized multi-word item); b) the degree of fixedness of a lexical combination is also measured in relation to its grammatical restrictions; c) finally, as non-compositionality is the basis of a collocational pattern, it is essential to assess, whether the meaning of a multi-word item can or cannot be obtained from the meaning of its constituent parts.

Corpora and data analysis

My first step was to design a body of non-technical English texts of late Middle English that could offer a **Non-Technical Corpus** (henceforward**, n-TC**) that might be activated as a corpus of reference. Secondly, I built up a minor body of the legal texts of the same period (**Le**gal **Co**rpus, henceforward **LC**). **n-TC** and **LC** were for the most part sketched according to two simple features: a) a "medium-oriented choice": the texts were basically selected according to their electronic readability. For this purpose, The Humanities Text Initiative, a unit of the University of Michigan's Digital Library Production Service, offered me online access to full text resources of the *Corpus of Middle English Prose and Verse*; and b) a wide-ranging "topic-oriented choice": **n-TC** texts were chosen according to their common (non-technical) character. Thus the type of text selected fits in a broad area of topics which might have embodied the common speech of the fifteenth century: fiction, drama and religious texts. Table 1 shows the texts of the reference corpus and the number of words in each:

LC[3] has been pictured on the basis of its restricted legal technical character and it has also been apportioned into seven subgenres: a) Depositions, b) Lincoln Documents, c) Gylds, d) Indentures, e) Petitions, f) Signet and Privy Seals, and g) Wills.

Table 2 shows the data of LC including the words of subgenres.

In addition to these internal features of the corpora, **n-TC** and **LC** show different external contextual characteristics. **n-TC** embodies extensive 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.),

TEXTS OF N-TC	WORDS
The Canterbury Tales	206,734
Everyman	8,118
Confessio Amantis	241,707
Orpheus and Eurydice	27,703
The minor Poems of Robert Henryson	5,392
The morall Fabillis of Esope the Phrygian	5,346
The testament of Cresseid	7,016
The Vision of Piers Plowman	79,767
Octavian	11,627
Pearl	8,418
Sir Gawain and the Green Knight	21,270
The Alliterative Morte Arthure	42,112
The Siege of Jerusalem	12,044
The Towneley Plays	99,943
The York Plays	99,736
Troilus	75,014
Rewle	15,736
Merlin	293,332
	1,261,015

TABLE 1: n-TC texts

SECTION/SUBGENRE	WORDS	%
Depositions	2,067	0.50
Documents	75,589	18.50
Gylds	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

TABLE 2: LC wordlists of subgenres

different types of text (drama, correspondence, fiction, history, romance, etc) and different topics (religion, fiction, etc.). However, **LC** external contextual features are more limiting. Thus, legal texts are functionally informative and exhibit both a statutory style and a formal/professional and distant down setting.

The Wordlist programme (Wordsmith) reckoned 1,261,015 tokens (running words) for the total word estimate of **n-TC**, whereas for **LC** the programme computed 408,398 instances. The wordlists include tokens and rates, but it was necessary to detect accurately "types" of tokens/words. For the sake of simplicity I have included under the same type both the spelling and morphological forms of a word. The programme has computed 67,155 types for n-TC and 26,032 for LC.

N^1 + N^2 collocational type

Now I will initiate the analysis of the collocational binary pattern N^1 + N^2 as in *lawe day*. Functionally, such a structure is a collocation in which N^2 operates as head (node of collocation) and N^1 as a premodifier or left-collocate rather than an Anglo-Saxon genitive, as no possession is implied. My analysis focuses on those salient lexical words which occur in LC as nodes (N^2). I take for granted that the more salient a word is in LC, the more likely it is to form collocational patterns. To restrict my research, the present paper is confined only to the top 24 LC salient N^2 detected in the first 100 saliency rank scale and exhibited in Table 3.

It is important to point out that saliency is attained by comparing the frequency rates of both LC and n-TC. The more frequent a word is in LC and the less common in n-TC, the more salient it is in LC. Thus, the most frequent noun word in LC is *day* with 801 tokens, however the noun *yere* is the most salient word in LC with 649 occurrences because by comparing the number of tokens of these two words in n-TC we observe that *day* appears 2,082 times, whereas *yere* occurs only on 257 occasions. Thus the programme includes *yere* in first position of saliency, whereas *day*, which is the most frequent noun in LC, occurs in the tenth position in the saliency rank scale. Although I assume that saliency should play an important role in a collocational framework, only the seven words displayed in bold in Table 3 show N^1 + N^2 collocational patterns.

The first salient N^2 in LC that exhibits the collocational pattern N^1 + N^2 is *chirche* (546 occ.). The search conducted for all concordances provides 64 examples of *chirche* as N^2. Only 4 types of N^1 collocate with the node *chirche* with 4 tokens or more: *parische* (LC 28 occ., n-TC 2), *cathedral* (LC 18 occ, n-TC 1), *moder* (LC 10 occ., n-TC 1) and *cryst(s)* (LC 4 occ., n-TC 1). Consider examples (1) and (2):

LC RANK	NOUN WORD SALIENCY	LEXICAL WORD	FREQ.LC	%	FREQ. N-TC	%	KEYNESS	*P*-SCORE
24	1	YERE	649	0.49	257	0,02	**2,269.1**	0,000000
25	2	**CHIRCHE**	546	0.41	163	0,01	**2,044.6**	0,000000
30	3	ALDERMAN	356	0.27	3		**1,779.6**	0,000000
32	4	SOULE	402	0.31	58		**1,709.2**	0,000000
43	5	EXECUTOR	281	0.21	7		**1,366.8**	0,000000
44	6	FRATERNITE	266	0.20	2		**1,331.9**	0,000000
53	8	**KYNG**	216	0.16	10		**1,020.0**	0,000000
55	10	**DAY**	801	0.61	2.082	0,13	**1,014.6**	0,000000
60	11	PARTE	318	0.24	195	0,01	**970.5**	0,000000
63	12	LORD	203	0.15	13		**937.9**	0,000000
64	13	**TYME**	662	0.50	1.550	0,10	**926.5**	0,000000
65	14	CYTE	283	0.21	141		**925.4**	0,000000
70	15	TOWNE	263	0.20	119		**885.3**	0,000000
75	16	**TERME**	252	0.19	113		**850.5**	0,000000
76	17	TENEMENT	176	0.13	7		**838.2**	0,000000
80	18	TESTAMENT	201	0.15	41		**810.4**	0,000000
81	19	SUCCESSOUR	159	0.12	0		**809.9**	0,000000
86	20	HEIRES	190	0.14	45		**745.6**	0,000000
96	21	**CLERC**	174	0.13	45		**671.2**	0,000000
97	22	MANER	188	0.14	71		**665.0**	0,000000
98	23	BODY	222	0.17	144		**663.7**	0,000000
100	24	**CHARGE**	213	0.16	144		**627.0**	0,00000

TABLE 3: Salient noun words in LC

1. Also, I bequeth vnto the parishe chirche of Leylond to bye a grette bell to tenour those iiij. other belles (*Lincoln*)
2. Also, I bequeth to the mother chirche of Lincoln (*Lincoln*)

The next salient N^2 involved in a collocational pattern $N^1 + N^2$ in LC is *Kyng* (417 occ.) whose concordance search shows that only *lord* (LC 14 occ., n-TC 5) and

prince (LC 6 occ., n-TC 0) occur as left-collocates with *kyng* as in (3) and (4). Consider

> 3. Henry the fourthe youre Aiell And to the full nobill and gracious prince kyng Henry þe fifte your fader whos soules god assoile to graunte. (*Chancery*)
> 4. in the xxvth yere of the reigne off our said souereign lorde kyng henry the eight, (*Lincoln*)

The search carried out for all concordances of *day*, next in the saliency rank scale, shows 10 collocational subtypes of N^1 + N^2 with 4 tokens or more. Consider (5) and (6):

> 5. but the dedication day of the church was (*Lincoln*)
> 6. to holde their lawe day in the Guyldehall (*Gylds*)

It is noteworthy that *day* as N^2 collocates with 21 different N^1 at least twice. Thus, *market* (12 occ., n-TC 0), *dedication* (7 occ., n-TC 0), *Gylde* (6 occ., n-TC 0), *esterne* (5 occ., n-TC 4), *lawe* (4 occ., n-TC 0), *eleccion* (4 occ., n-TC 0), *candilmesse* (4 occ., n-TC 3), *michelmas* (4 occ., n-TC 3), *halowen* (4 occ., n-TC 5), *lammesse* (4 occ., n-TC 5), *hallomese* (3 occ., n-TC 3) *cristemas* (3 occ., n-TC 7), *festiual* (3 occ., n-TC 0), *mydsomer* (3 occ., n-TC 2), *lady* (2 occ., n-TC 1), *weke* (2 occ., n-TC 1), *all saints* (2 occ., n-TC 3), *ascention* (2 occ., n-TC 4).

The N^2, *tyme,* is the next in the saliency rank and shows that 9 types of N^1 collocate with it in left position: *seruice* (4 occ.), *market* (4 occ.), *cession* (2 occ.), *terme* (2 occ.), *easter* (2 occ.), *feyre* (2 occ.), *winter* (1 occ.), *harvest* (1 occ.), *night* (1 occ.) as in (7) and (8):

> 7. to brenne in seruice tyme eueriche festiuale (*Gylds*)
> 8. all the markett tyme, in a whyte sheat (*Lincoln*).

The following salient node noun is *terme* and collocates with four types of N^1: *Easter* (6 occ.), *Trinity* (5 occ.), *Hillarie* (3 occ.), and *michaelmas* (3 occ.). Consider:

> 9. In Easter terme, vpon the ascention day; In Trinite terme, vpon the natyuy (*Lincoln*)
> 10. In Trinite terme, vpon the natyuyte of saincte Iohn baptist; In Michaelmas terme, vpon Alholon day; In Hillarie terme, vpon Candelmas day; (*Lincoln*)

The next salient N^2 is *clerk* which collocates with three types of N^1: *towne* (12 occ), *churche* (3 occ) and *parish* (1) as in (11), (12) and (13):

> 11. Baillyfs, to holde their fferyng day, and therupon the seide Toune clerk (*Gylds*)
> 12. I bequethe to þe Churche Clerk of Seynt Benet (*Wills*)
> 13. There were also the fees of the parish-priest, the parish-clerk (*Lincoln*)

Finally, in the saliency rank scale is *charge*, which includes as left-collocate the N1 *rent* (3 occ.) as in (14):

> 14. Wife to have her Dower in his inherited Lands and a Rent-Charge of £20 a-year (*Wills*)

N *and* N collocational pattern: types and tokens

Many twin collocations must have been formed when English made its way into the language of court proceedings, particularly after the *Statute of Pleading* (1362) when English was fostered as the language of local courts. In this transition process there was a great deal of hesitation about the exact meaning of the English words, so law experts tried to be self-confident by using the English and the French legal words together as in *landes and possessions, goods and chattels, will and testament, free and clear, right and interest, breaking and entering*, etc. In the course of time these *and*-nominal concordances were used to reinforce the meaning of a given form regardless of their provenance, as in *fourme and condition*.

All **N** *and* **N** concordances found in LC and attested to as collocations were contrasted with the figures of these collocational patterns in n-TC, so we may infer the acceptance of these technical collocations in other common registers. It is worth noting, however, that only those **N** *and* **N** concordances authenticated at least 6 times are included in LC. However, all instances are counted in n-TC because when an N *and* N-concordance has been verified as a collocation in the legal corpus, a single appearance in the other non-technical register allows us to believe that a specialised collocation has gone through other non-technical domains.

I should remark that my investigation concentrates exclusively on **N** *and* **N** collocations (eg. *landes and tenementes*) as in (15).

> 15. then I will that all the foresaide landes and tenementes to be equally deuyded emonges my children (*Lincoln*).

So no intervening elements are present either as determiners or as modifiers in the collocational structure. Thus the collocational type *þe Kyng and þe lords* (6 occ.) is not included in this category.

Table 4 includes 42 N *and* N collocational types and 791 tokens found in LC, and their distribution in the different subgenres and in n-TC. The data cover only those collocational patterns which include 6 occurrences or more in LC, although all instances are computed in the subcorpora.

Rank	Collocation	Type	Subgenres of LC							Total LC	Total n-TC
			Ind	Pet	Wil	Seal	Dep	Gyl	Doc		
1	Landes and tenements	N + N	1	15	23	3	1	12	52	**117**	0
2	Maister and wardens	N + N	0	0	0	0	0	80	0	**80**	0
3	Baillies and communes	N + N	0	0	0	0	0	60	0	**60**	0
4	Maner and fourme	N + N	0	10	9	2	0	22	17	**60**	0
5	Day and yere	N + N	2	3	6	0	1	3	19	**35**	0
6	Costes and charges	N + N	0	6	0	0	0	7	17	**30**	0
7	Testament and will	N + N	0	0	24	0	0	0	4	**28**	0
8	Fraternite and gyld	N + N	0	0	0	0	0	24	0	**24**	0
9	Maire and bailiffs	N + N	0	7	0	3	0	13	0	**23**	0
10	Executors and assignes	N + N	0	0	1	00	0	0	22	**23**	0
11	Issues and profites	N + N	0	3	6	0	0	1	11	**21**	0
12	Will and testament	N + N	0	0	2	0	0	0	17	**19**	0
13	Maister and brethern	N + N	0	0	0	0	0	16	0	**16**	0
14	Godes and catalles	N + N	0	3	4	1	0	5	2	**15**	0
15	Heiress and assignes	N + N	7	0	1	1	0	0	6	**15**	0
16	Rentes and profits	N + N	0	0	2	0	0	12	0	**14**	0
17	Alderman and maisters	N + N	0	0	0	0	0	13	0	**13**	0
18	Forme and condicions	N + N	0	0	0	0	0	0	12	**12**	0

Rank	Collocation	Type	Subgenres of LC							Total LC	Total n-TC
			Ind	Pet	Wil	Seal	Dep	Gyl	Doc		
19	Maire and sheriff	N + N	0	0	0	0	0	12	0	**12**	0
20	Town and marches	N + N	0	0	0	12	0	0	0	**12**	0
21	Felowes and scolers	N + N	0	0	0	0	0	0	10	**10**	0
22	Prior and chanons	N + N	0	8	0	0	0	2	0	**10**	0
23	Fraternite and crafte	N + N	0	0	0	0	0	9	0	**9**	0
24	Liberte and franchises	N + N	0	5	0	3	0	1	0	**9**	0
25	Wevers and fullers	N + N	0	4	0	5	0	0	0	**9**	0
26	Dean and chapter	N + N	0	0	0	0	0	4	4	**8**	0
27	Ende and terme	N + N	1	0	0	0	0	0	7	**8**	0
28	Name and fame	N + N	0	0	0	0	0	7	1	**8**	0
29	Autorite and power	N + N	0	2	0	1	0	4	0	**7**	0
30	Maner and tenements	N + N	5	0	0	0	0	0	2	**7**	0
31	Proffites and comodities	N + N	0	4	0	0	0	0	3	**7**	0
32	Proffites and reuenues	N + N	0	5	1	0	0	1	0	**7**	0
33	Wardons and felowship	N + N	0	0	0	0	0	7	0	**7**	0
34	Rentes and seruices	N + N	0	1	3	0	0	2	1	**7**	0
35	Tenementes and rentes	N + N	0	1	3	0	0	3	0	**7**	0
36	Chaplain and keeper	N + N	0	0	0	0	0	0	6	**6**	0

Rank	Collocation	Type	Subgenres of LC							Total LC	Total n-TC
			Ind	Pet	Wil	Seal	Dep	Gyl	Doc		
37	Heiress and executours	N + N	3	0	0	0	0	0	3	**6**	0
38	Maister and merchant	N + N	0	6	0	0	0	0	0	**6**	0
39	Priories and possesions	N + N	0	0	0	6	0	0	0	**6**	0
40	Rentes and fermes	N + N	0	0	0	3	0	0	3	**6**	0
41	Tenementes and heredyt	N + N	0	0	0	0	0	0	6	**6**	0
42	Day and place	N + N	0	2	0	0	0	3	1	**6**	1
	Total tokens		19	85	85	40	2	323	226	**791**	1

(Subgenre abbreviations: Ind = Indenture; Pet = Petitions; Wil = Wills; Seal = Signet and Privy Seals; Dep = Depositions; Gyl = Gylds; Doc = Documents)

TABLE 4: N *and* N collocational types and token

The examination of N *and* N-collocational types and tokens in the 7 subgenres (cf. Table 4) analysed in this research reveals an uneven distribution. Thus 14 out of 42 collocational types (33.3%) appear in a single subcorpus. For example, *maister and wardens*, which is the second most recurrent LC N *and* N-collocational type with 80 tokens, is detected only in the *Gyld* subcorpus (cf. Table 4). It is also important to note that 13 types (30.95%) are found only in 2 subgenres. Thus, for example, *testament and will*, occurs only in *Wills* (24 occ.) and *Lincoln Documents* (4 occ.) subgenres. 7 types occur in 3 subgenres such as *costes and charges* and only 3 types such as *issues and profites* are distributed in 4 subgenres. Similarly, only 3 patterns such as *maner and fourme* are detectable in 5 subgenres, and a single type, *day and yere*, is located in 6 subcorpora. *Landes and tenements*, however, is detected in all subgenres and it is very recurrent as it accounts for 117 collocations.

The most outstanding feature is the absence of these N *and* N-collocational patterns in the common corpus as only a single instance of *day and place* is found in n-TC, which ascertains that they were restricted to legal English.

ADJ + N collocational type

A survey of adjectives either as nodes or collocates shows a significant feature worth mentioning: by checking the 1,000 most common words in LC I have only found 26 different 'qualifying' adjectives. The relatively infrequent use of adjectives is an expected feature of technical legal language. It is also assumed that when they occur they are likely to be recurrent. In fact, 22 out of the 26 adjectives found in the 1,000 most common words in LC occur at least 36 or more times.

The keyword programme shows, however, that 14 of these adjectives were even more frequent in non-technical corpus such as *gret, right, good,* etc. In fact, only the 8 adjectives displayed in Table 5 are salient in LC. Notwithstanding, their saliency is so great that it reveals that they were restrictedly used in technical legal structures forming collocational patterns with a quasi-formulaic connotation.

LC RANK	SALIENCY RANK	ADJECTIVE	FREQ.LC	%	FREQ. N-TC	%	KEYNESS	*P*-SCORE
226	49	souerain	184	0.05	2		526.0	0.000000
238	55	welbeloued	161	0.05	2		434.0	0.000000
241	69	trusty	145	0.05	15		277.0	0.000000
267	116	worshipful	132	0.04	8		207.3	0.000000
311	261	gracious	113	0.02	34		121.5	0.000000
215	282	holy	215	0.06	345	0.02	115.6	0.000000
835	309	discrete	41	0.01	7		106.0	0.000000
687	458	laufull	72	0.01	12		78.5	0.000000

TABLE 5: LC salient adjectival types for the first more frequent 1,000 words

Now I proceed with a detailed account of LC salient adjectives, specifying their collocational patterns:

1. *Souerain* (LC rank 226; Salient word rank 49; Keyness = 526.0; $p = 0.000000$)

Souerain is the first LC salient adjective, though it appears in 49th position of all salient words. For the purpose of this work I have run the concord programme to catch a glimpse of collocational patterns with *souerain*. I have detected that it is recurrently used in expressions such as *the kyng oure souerain Lorde*. The programme has spotted 184 tokens of *souverain*, most of them functioning as right collocates of *kyng* and left-collocates of *lord*. Note that all instances were found in *Chancery*. Examples include:

16. To the kyng oure souerain lord Right mekely besecheth to your souerain lordship youre pore seruant William (*Chancery*).
17. Please it to the Kyng oure souerain Lord of youre Benigne grace to graunte to youre humble seruants (*Chancery*).

On the basis of the data provided by the programme we must conclude that *souerain* was associated with *kyng* which, in syntactical terms, must be considered and forming with *lord* a post-complementation of *kyng* or an appositive structure.

2. *Wellbeloued* (LC rank 238; Salient word rank 55; Keyness = 434.0; p = 0.000000)

Although *wellbeloued* appears with far from top saliency (55th position), it is the second most salient adjective in LC not only because it includes 161 tokens, but also because it is very infrequent in n-TC. Furthermore *wellbeloued* is a genre-dependent adjective as its usage is mainly restricted to *Chancery* with 159 out of 161 tokens (*Lincoln* 1, *Wills* 1). The concord programme has found the collocational type *oure right trusty and welbeloued* 67 times as in (18) and (19):

18. Signet of Henry V By þe king Worshipful fader in god / right trusty and welbeloued / we grete yow wel / (*Chancery*).
19. Signet of Henry V By þe king Worshipful fader yn god. oure right trusty and welbeloued. We grete yow wel (*Chancery*).

The cluster *oure trusty and welbeloued* (*lord, kyng, brother, cousin, wife, clerc, squire*) is also found on 66 occassions as in (20) and (21):

20. Signet of Henry V By the king Trusty and welbeloued We grete you wel (*Chancery*).
21. We haue receyued a supplicacion put vn to vs be our trusty and welbeloued knyght henry Brounflete that is with vs in our seruice (*Chancery*).

Wellbeloved may be also pre-modified by *entirely* in these collocational patterns as in (22)

22. Right trusty & entierly welbeloued frende / I grete you often tymes wel / And thanke you wit (*Chancery*).

It is worth noting that 32 occurrences of *wellbeloved* are also found qualifying a personal noun (*wife, clerc, cousin*, etc.) without being clustered in a collocational pattern. Consider (23):

23. vnder oure signet contenyng certain articles aduised by oure welbeloued knyht Iohn Tiptoft Seneschal of oure duchie of guyenne (*Gylds*).

3. *Trusty* (LC rank 241; Salient word rank 69; Keyness = 277.0; p = 0.000000)

Trusty appears in third position in the adjective saliency list and occupies the 69th place in LC rank with 145 tokens. It exhibits three collocational types:

i) *trusty and welbeloued* (*lord, kyng, frend, cousin, brother*, etc.), which includes 136 tokens and is mainly restricted to *Chancery* (134 occurrences, *Wills* 2). Consider (24):

> 24. Iohn Bisshop of Saint Asseph. William Bisshop of Salesbury and oure trusty and welbeloued Cousyn William Erl of Suffolk (*Chancery*).

Trusty is frequently premodified by *right* as in (25):

> 25. Signet of Henry V By þe kyng Worshipful fader in god right trusty and welbeloued / We grete yow wel (*Chancery*).

ii) *Dere and trusty* (two tokens, both in *Wills*) as in (26):

> 26. bot þe holy trinite kepe 3ow now, dere and trusty wyf? (*Wills*).

iii) *Effectuel and trusty* (two tokens, both in *Chancery*) as in (27):

> 27. Lady of wilton: in the whiche matire y praie yow. that ye be effectuel and trusty frend (*Chancery*).

It is worth remarking that just a single instance out of 141 is found as a free qualifying adjective as in (28):

> 28. I ordeyn my trusty frendes, Iankyn? Miles, Thomas Knolles aforsaid?, Elizabet? Ioy (*Wills*)

4. *Worshipful* (LC rank 267; Salient word rank 116; Keyness = 207.3; $p = 0.000000$)

Worshipful continues as the 4th most LC-salient adjective and fills the slot 116 in the saliency rank. It includes 132 tokens. It is basically restricted to the collocational type *Worshipful fader* (131 instances) within the formulaic expression *Worshipful fader yn god ri3t trusty and welbeloued*, and it is also domain-dependent as 128 tokens appear in *Chancery*. The other 3 occur in *Lincoln*. Consider (29) and (30)

> 29. Signet of Henry V Worshipful fader yn god ri3t trusty and welbeloued / We grete yow wel / (*Chancery*).
> 30. To the worshipful and wyse syres and wyse Communes that to this present (*Chancery*).

This collocational framework in LC has only 8 instances in the common corpus.

5. *Gracious* (LC rank 311; Salient word rank 261; Keyness = 121.5; $p = 0.000000$)

The next LC-salient adjective is *gracious,* which appears in slot 261 with regard to LC rank with 113 occurrences. It shows various collocational types: *your gracious*

lordship (24 occ.), *your noble and gracious* + N (10 occ.), *your worthy and gracious* + N (7 occ.), *your gracious lettres patentes* (7 occ.) Consider (31) and (32):

> 31. Plese it to your gracious lordship to considre þe premisses and þer uppon to graunte (*Chancery*).
> 32. to aduertice that howe oure souerain lord by his gracious lettres patentes made vndir his grete seal of Englond (*Chancery*).

In most instances (61) *gracious* premodifies a head noun without forming a collocational pattern as in (33):

> 33. but we haue the sounere remedie by youer most gracious socour and helpe at this present parlement (*Chancery*).

Gracious is mostly used in *Chancery* (97 tokens), but it also appears in the three other texts (*Lincoln* 9, *Gylds* 4, *Wills* 3)

6. *Holy* (LC rank 215; Salient word rank 282; Keyness = 155.6; p = 0.000000)

Holy comes next in the saliency rank scale for adjectives (6th position) and it is exhibited in 215th position in LC rank with 215 occurrences. It shows various collocational types: *holy chirche* (32 occ.), *holy company* (17 occ.), *holy gost* (16 occ.), *holy rood* (13 occ.), *holy cross* (13 occ.), *holy trinity* (12 occ.), *holy martyr* (11 occ.), *holy father* (8 occ.) and *holy days* (6 occ.) as in (34) and (35):

> 34. In þe worchep of God of heuen, and of his modir seynt mari, and alle þe holy Company of heuen, and souerengly of þe Noble confessour (*Gylds*).
> 35. In the name of the fadir and the sone and the holy goste, I, Isabelle Maryone, of your diocese, wydowe, behest (*Lincoln*).

Holy appears in all legal subgenres and is also frequent in the common corpus with 345 tokens.

7. *Discrete* (LC rank 835; Salient word rank 309; Keyness = 106.0; p = 0.000000)

Discrete appears in 7th place in the LC-salient adjective list and fills slot 835 in LC rank with 41 tokens. *Discrete* was only used in two legal subgenres (*Chancery* 28 occ. and *Gylds* 13 occ.). It shows two basic collocational patterns: *discrete* + N and Adj *and discrete* + N displayed in the following types: i) *discrete councel* (23 occ.), *discrete comunes* (10), *discrete persones* (9 occ.); and ii) *wyse and discrete* + N (9 occ.), *high and discrete* + N (6 occ.), *sadd and discrete* + N (5 occ.), *worshipful and discrete* + N (4 occ.) as in (36) and (37):

> 36. Petition of James, Earl of Wiltshire To the full wyse and discrete Comons of this present parlement (*Chancery*).
> 37. to þe hygh and discrete councell of oure souueraign lord (*Chancery*).

8. *Laufull* (LC rank 687; Salient word rank 458; Keyness = 78.5; p = 0.000000)

The 8th most salient adjective, *lauful*, goes to 687th position in LC rank with 72 tokens, and there are 457 words which are more salient in LC. It is profusely employed in *Lincoln* subgenre with 51 occurrences, only 13 were detected in *Gylds*, 7 in *Chancery* and 1 in *Wills*. *Laufull* occurs basically in 2 different collocational patterns displayed as follows: *laufull* + (*English*) *money* (12 occ.), *laufull* + *impediment* (6 occ.) and *trewe, just, and laufull men* (5occ.) as in (38) and (39):

> 38. her be spent att my buryall and att my monethes daie fifty poundes of laufull money off Englond (*Lincoln*).
>
> 39. the soules aboue remembred, and for all christen soules, hauyng noo laufull impedyment; (*Lincoln*).

To conclude this section we may affirm that salient adjectives are domain dependent as all of them, except *holy*, are predominantly restricted to a subgenre. Now I will proceed to analyse whether a salient noun is likely to be qualified, and if so, recurrently qualified by the same adjective so as to form a collocation. A review of the 24 salient nouns (see Table 3) which occur in the 100 salient words shows that only 10 are commonly modified by the same adjective, but it is also important to note that only 5 salient adjectives (*holy, souerain, worshipfull, trusty* and *wellbeloued*) combine with salient nouns. It is also frequently the case that the same adjective collocates with different nodes. Thus *holy* is a left-collocate of *yere*, *chirche* and *day*. *Souerain* is a left-collocate of *kyng* and *lord*. *Worshipful* also collocates in left position with *father, kyng*, *lord* and *town*. *Trusty* and *Wellbeloued* are right and left-collocates of *lord, kyng and clerc*.

Adj *and* Adj + N collocational pattern: types and tokens

Similarly to twin *and*-nominal collocations I have also detected a large number of twin adjectival *and*-concordances such as *trusti and wellbeloued* which had great saliency in LC. Table 6 shows types and tokens through the subgenres:

The collocational pattern Adj *and* Adj + N contains 14 types and 262 tokens. The most frequent type is *trusti and wellbeloued*, which is found on 131 occasions in LC. It is worth mentioning that *trusty* appears as a salient word in LC in 61st rank position, but it turns up in first position when it is calculated as forming the collocation *trusti and wellbeloued*. It is also significant that all instances were registered within the sequence *(right) trusty and welbeloued* as in (40):

> 40. By þe king Worshipful fader yn god right trusty and welbeloued. We grete yow wel (*Chancery*).

Rank	Collocation	Type	Subgenres of LC							Total LC	Total n-TC
			Ind	Pet	Wil	Seal	Dep	Gyl	Doc		
1	Trusti & wellbeloued	Adj + Adj	0	0	0	131	0	0	0	**131**	0
2	Spiritual & temporel	Adj + Adj	0	31	0	3	0	3	2	**41**	0
3	God & lawful	Adj + Adj		0	0	0	0	7	7	**14**	0
4	Content & paid	Adj + Adj	1	0	0	0	0	0	12	**13**	0
5	Wise & discret	Adj + Adj	0	10	0	1	0	0	0	**11**	0
6	Gode & trewe	Adj + Adj	0	2	1	0	0	6	0	**8**	3
7	Ferme & stable	Adj + Adj	0	0	0	0	0	7	0	**7**	2
8	Complet & ended	Adj + Adj	0	0	0	0	0	0	6	**6**	0
9	Right & lawful	Adj + Adj	0	0	0	0	0	6	0	**6**	0
10	Used & accustomed	Adj + Adj	0	0	0	2	0	4	0	**6**	0
11	Gode & effectuel	Adj + Adj	0	5	0	0	0	0	0	**5**	0
12	Grete & notable	Adj + Adj	0	5	0	0	0	0	0	**5**	0
13	Noble & gracious	Adj + Adj	0	5	0	0	0	0	0	**5**	0
14	Worthy & gracious	Adj + Adj	0	4	0	0	0	0	0	**4**	0
	Total tokens		1	62	1	137	0	33	27	**262**	5

TABLE 6: Adj *and* Adj + N collocational types and tokens

The examination of Adj *and* Adj collocational types and tokens in the 7 subgenres (cf. Table 6) surveyed in this research reveals an unequal distribution. Thus 8 out of 14 collocational types appear in a single subcorpus. For example, *trusti and wellbeloued*, which is the most recurrent LC Adj *and* Adj-collocational type with

131 tokens, is detected only in the *Seal* subcorpus. It is also important to note that 4 types are found only in 2 subgenres. Thus, for example, *god and lawful*, occurs only in *Gylds* and *Lincoln Documents* subgenres. Only 1 type is found in 3 subgenres, *gode and trewe,* and also a single type, *spiritual and temporal*, is distributed in 4 subgenres. No instances were detected in all subcorpora.

The most notable aspect is the absence of this Adj *and* Adj-collocational pattern in the common corpus as only two types *gode and trewe* (3 occ.) and *ferme and stable* (2 occ.), are found in n-TC, which indicates that they were predominantly restricted to legal English.

V + N collocational pattern

Collocational patterns such as *take charge*, usually known as composite predicates, have been profusely studied since Jespersen (1942: 117), who considers the support verb of the pattern as a "lexically empty verb". Basically the structure includes two main features: on the one hand, a support verb with a very restrictive telic significance, but provided with all grammatical features, such as the inflections to mark tense, number and person (always animate), and on the other hand, a deverbal abstract noun which functions like a verb as it carries the action and the lexical meaning. Many authors have studied composite predicates in modern English (Live 1973; Björkman 1978; Gross: 1986; Pivaut 1994). There is also a specifically historical overview of composite predicates in the history of English in Brinton and Akimoto (1999).

Composite predicates come from OE, though they were restricted to the verbs (*ge*)*don*, (*ge*)*macian, sellan, giefan, niman* and *habban* (Brinton and Akimoto: 1999: 21-58). However, Matsumoto (1999:59-95) confirms that composite predicates were extensively used from the 13th century onwards. The widespread use of composite predicates in the course of the 14th and 15th centuries was probably reinforced by similar patterns in French (Iglesias-Rábade: 2000: 93-130).

A survey of collocational composite predicates in my corpora shows that the verbs *taken, giuen, hauen* and *beren* are often used as predicators with little telic significance as the meaning has been relocated to the following deverbal abstract noun as in (41):

> 41. the whiche I haue long taught, holden, and yeven faith & credence to theym, agayn many and diuers holy sacramente (*Lincoln*).

Table 7 shows V + N collocational patterns with the verbs *taken, giuen, hauen* and *beren* in LC and n-TC:

SUPORT-VERB	DEVERBAL NOUN	LC-TOKENS	N-TC TOKENS
Take	charge	15	5
	advauntage	4	2
	profit	3	7
	suit	3	0
	hede	3	61
	action	3	0
Geue	assent	6	0
	licence	4	0
	grace	4	3
	verdict	3	0
	evidence	3	0
	warning	3	0
	faith and credence	3	0
	notice	3	0
Haue	grace	23	9
	power	7	10
	mercy	7	8
	auctorite	4	0
	cause	4	12
	knowledge	3	59
Bere	charges	4	0
	office	4	0
	witness	3	36
	armes	3	0
	costes	3	0

TABLE 7: V + N collocational patterns

The n-TC includes many types of collocational composite predicates with *take*, apart from those found in LC. Thus I found the following types with five tokens or more: *take vengeance* (13 occ.), *take armes* (12 occ.), *take counseil* (12 occ.), *take rest* (8 occ.), *take evidence* (8 occ.), *take efeect* (7 occ.), *take witnesse* (5 occ.), *take journey* (5 occ.), *take avis* (5 occ.), *take heart* (5 occ.), *take hold* (5 occ.), etc. I have not detected the types found in the LC in the n-TC with the support verb *geue*, except *geue grace*. With regard to *haue*, all types found in LC were also detected in n-TC, except *have auctorite*. It is also worth noting that the n-TC includes many other types that have not been detected in the LC such as *haue pity* (13 occ.), *haue pees* (9 occ.), *haue doute* (8 occ.), *haue shame* (7 ooc.), *haue succour* (7 occ.), *haue blame* (6 occ.), *haue honour* (5 occ.), etc. The support verb *beren* has no collocational support verb + deverbal noun in n-TC, except *beren witnesse*, which includes 36 instances, but almost all of them in *Piers Plowman* with a formulaic character as in (42):

42. Was that Sarsen saved, as Seint Gregorie bereth witnesse. Wel oughte ye lordes that lawes kepe this lesson (*Piers Plowman*: 11.156-157).

Collocational composite predicates are more common in n-TC than in LC with the support verbs *take* and *haue*. Except *take suit, take action* and *haue auctorite*, all types of composite predicates found in the legal corpus were also found in n-TC. Furthermore, these two support verbs were very productive, forming a varied range of types of composite predicates in non-technical English.

V *and* V collocational pattern

Likewise to twin *and*-nominal and adjectival collocations, my corpora also show an extensive use of twin verbal *and*-concordances such as *ordeyne and make* which exhibit practically an exclusive presence in LC. Table 8 shows types and tokens through subgenres.

As is shown in Table 8, the legal corpus exhibits 22 types with 6 or more occurrences that incorporate 279 tokens in the category V *and* V. *Couenaute and graunte* with 32 tokens is the most recurrent *and*-concordance in this class and it appears in 8th position in the collocation type rank as in (43):

43. Firste, the seid Thomas Tanfelde covenantes and grauntes, by thies presentes, for to cause a yer (*Gyilds*).

An examination of V *and* V collocational types and tokens in the 7 subgenres (cf. Table 8) checked in the corpora exhibits an unequal distribution, although most collocational types are used in more than one subgenre. Only 4 types (*haue and hold, liberate and allocate, yeld and pay* and *occupy and inioye*) appear in a single subcorpus, although recursively used. For example, *haue and hold* is detected only in the *Lincoln Documents* subcorpus, but it occurs 23 times. It is also important to note that 12 types are found only in 2 subgenres. Thus, for example, *ordeyne and dispose* occurs only in the *Wills* and *Lincoln Documents* subgenres. Similarly 3 types, including *geue and bequeth*, are found in 3 subgenres, whereas only two types, such as *ordeyne and assigne*, are exhibited in 4 subcorpora. Likewise, 2 types were extensively used in 5 subgenres, as in *make and ordeyne*, but no instances were detected in all subcorpora.

It is noticeable that no V *and* V-collocational pattern of those shown in Table 8 occurs in n-TC, except a single type, *pray and require*. Conclusively V *and* V concordances were also predominantly restricted to legal English.

Rank	Collocation	Type	Subgenres of LC							Total LC	Total n-TC
			Ind	Pet	Wil	Seal	Dep	Gyl	Doc		
1	couenaute and graunte	V + V	0	0	0	0	0	9	24	**32**	0
2	ordeyne and make	V + V	0	2	3	1	0	3	15	**24**	0
3	make and ordeyne	V + V	0	1	10	1	0	8	3	**23**	0
4	haue and hold	V + V	0	0	0	0	0	0	23	**23**	0
5	dirige and masse	V + V	0	0	0	0	0	8	12	**22**	0
6	geue and bequeth	V + V	0	0	1	0	0	1	15	**17**	0
7	liberate and allocate	V + V	0	0	0	12	0	0	0	**12**	0
8	yeld and pay	V + V	0	0	0	0	0	0	12	**12**	0
9	comen and offeren	V + V	0	0	0	0	0	9	2	**11**	0
10	ordeyne and establishe	V + V	0	6	0	0	0	4	2	**11**	0
11	entre and distraine	V + V	0	1	0	0	0	0	8	**9**	0
12	Obserue and kepe	V + V	0	0	0	0	0	1	8	**9**	0
13	ordeyne and dispose	V + V	0	0	4	0	0	0	5	**9**	0
14	singe and prey	V + V	0	0	6	0	0	2	1	**9**	0
15	haue and occupy	V + V	0	6	0	2	0	0	0	**8**	0
16	assent and consent	V + V	0	0	0	0	0	4	4	**8**	0
17	ordeyne and graunt	V + V	0	4	1	1	0	2	0	**8**	0
18	maintain and sustain	V + V	0	0	0	0	0	3	4	**7**	0
19	pray and require	V + V	0	2	0	0	0	0	5	**7**	5
20	couenaute and agree	V + V	0	0	0	0	0	2	4	**6**	0
21	occupy and inioye	V + V	0	0	0	0	0	0	6	**6**	0
22	ordeyne and assigne	V + V	0	2	0	1	0	2	1	**6**	0
	Total tokens		0	24	26	18	0	58	155	**279**	5

TABLE 8: V *and* V collocational types and tokens

Conclusions

The collocational patterns which have been retrieved and filtered in this study were categorised according to the semantic criteria of non-compositionality (meaning is not interpreted on a word-by-word basis), institutionalization (multi-word items are assumed to have a conventionalized value) and fixedness (item usage is subjected to grammatical restrictions).

As for the first collocational type, $N^1 + N^2$, I have compared LC and n-TC wordlists to check word saliency in LC, as I started out from the hypothesis that the salient words of the legal corpus were presumed to be prone to forming collocational patterns. However, this assumption turned out to be partially true under this category, as a large number of LC salient words were also very frequent in n-TC (e.g. *yere, chirche, kyng, day, lord, city*). The second category, the N *and* N collocational pattern, shows on the one hand an uneven distribution through legal genres. Thus 14 (33.3%) out of 42 collocational types appear in a single subcorpus, whereas only one type, *landes and tenements*, is detected in all subgenres. On the other hand an important aspect is the absence of N *and* N-collocational patterns in the non-technical corpus. Only a single instance of *day and place* is found in n-TC. The third category, the ADJ + N collocational pattern, shows a small number of types because the survey of the 1,000 most frequent words in LC exhibits only 26 different 'qualifying' adjectives. However, those which occur are very recurrent. Thus, 22 out of the 26 adjectives found in the 1,000 most common words in LC occur 36 or more times. In the category Adj *and* Adj I have detected a large number of types, although unevenly distributed across genres. Thus 8 out of 14 collocational types appear in a single subcorpus and no instances were detected in all subgenres. It is, however, interesting to observe that these collocational types are scarcely found in n-TC, which confirms that they were mostly restricted to legal English. Study of the collocational pattern V + N shows that only the verbs *taken, giuen, hauen* and *beren* are significantly used with no telic significance. I have found that the support verbs *take* and *haue* were exceptionally productive for this category in n-TC forming a varied range of types. Finally, the category V *and* V displays an extensive use, mostly also restricted to legal language. Thus only 1 out of the 22 types found in LC is recorded in n-TC. Their allocation to subgenres is very unbalanced. For example, 4 types appear in a single subcorpus, although recurrently used, and no instances were detected in all subgenres.

In broad terms the most important conclusion is that most of the collocational types detected in the LC in the six categories covered by this study were predominantly restricted to legal English.

Notes

[1]. This research has been funded by the Spanish Ministry of Education and Science, grant number HUM2005-00562/FILO and by La Dirección Xeral de I + D da Consellería de Innovación e Industria da Xunta de Galicia, grant number PGIDIT06PXIC204032PN. These grants are hereby gratefully acknowledged.

[2]. If two items *x* and *y* have probabilities of occurrence *p(x)* and *p(y)*, their mutual information *MI(x,y)* is formulated as

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

When $p(x,y) = p(x) / p(y)$ and the resulting value of *MI(x,y)* is 0, we may assert that the concurrent appearance of *x* and *y* is 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.

[3]. LC texts include four subcorpora:

a) *English Gylds. The Original Ordinances of more than 100 early English Gylds from 14th and 15th centuries.*
b) *An Anthology of Chancery English.*
c) *Lincoln Diocese Documents, 1450-1544.*
d) *Fifty earliest English Wills in the Court of Probate.*

Works cited

BACKUS, A. 2003. "Units in code-switching: evidence for multimorphemic elements in the lexicon", *Linguistics*, 14/1: 83-132.

BAKER M., G. FRANCIS and E. TOGNINI-BONELLI (eds). 1993. *Text and Technology*. Amsterdam: John Benjamins.

BALDWIN, T. 2005. "Looking for prepositional verbs in corpus data" In Blez, A. and S. Varges (eds.) *Proceedings of the 2nd ACL-SIGSEM Workshop on Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*. Colchester, UK:115-126.

BAZELL, C. E., J. C. CATFORD, M. A. K. HALLIDAY, and R. H. ROBINS (eds). 1966. *In Memory of F. R. Firth*. London: Longman.

BENSON, M. 1990. "Collocations and general-purpose dictionaries." *International Journal of Lexicography*, 3(1): 23-34.

BJÖRKMAN, S. 1978. *Le type avoir besoin. Étude sur la coalescence verbo-nominale en français*. Uppsala: Borgstöms Tryceri.

BRINTON, L. J. and M. AKIMOTO. 1999. *Collocational and Idiomatic Aspects of Composite Predicates in the History of English*. Amsterdam: John Benjamins.

CARTER, R. 1987. *Vocabulary: Applied Linguistic Perspectives*. London: Routledge.

CERMÁK, F. 2001. "Substance of idioms: perennial problems, lack of data or theory?",

International Journal of Lexicography, 14/1: 1-20.

CHOUEKA, Y. 1988. "Looking for needles in a haystack". *Proceedings of Recherche d'Information Assistée par Ordinateur* 1988. Cambridge, Mass.: 609-623.

CHURCH, K. and P. HANKS. 1989. "Word association norms, Mutual Information and Lexicography". *Computational Linguistics* 16(1): 22-29.

CLEAR, J. 1993. "From Firth principles. Collocational tools for the study of collocation." In Baker, M. et al. (eds) *Text and Technology.* Amsterdam: John Benjamins: 271-292.

COULTHARD, M. 1994 (ed.). *Advances in Written Text Analysis*. London: Routledge.

GITSAKI, C. 1996. *The Development of ESL Collocational Knowledge*. Unpublished PhD. thesis. Brisbane, Australia: The University of Queensland.

GLEDHILL, C. J. 2000. *Collocation in Science Writing*. Tübingen: Gunter Narr Verlag.

GROSS, G. 1986. "Les nominalizations d'expressions figées". *Langue Française*, 69: 64-84.

GROSSMANN, F. and A. TUTIN (eds.) 2003. *Les collocations: analyse et traitement.* Amsterdam: De Werelt.

HALLIDAY, M.A.K. 1961. *Categories of the Theory of Grammar*. Edinburgh: Edinburgh University Press.

—. 1966. "Lexis as a Linguistic Level". In Bazell, C. E., J. C. Catford, M. A. K. Halliday, and R. H. Robins (eds.) *In Memory of F.R. Firth*. London: Longman: 151-157.

IGLESIAS-RÁBADE, L. 2000. "The French phrasal power in late Middle English". In Trotter, D.A. (ed.) *Multilingualism in Later Medieval Britain.* Rochester: Boydell and Brewer: 93-130

—. 2001. "Composite predicates in Middle English with the verbs *nimen* and *taken*". *Studia Neophilologica*, 73: 143-163.

INKPEN, D. and G. HIRST. 2006. "Building and using a lexical knowledge base of near-synonym differences". *Computational Linguistics,* 32(2): 223-262.

JESPERSEN, O. 1942. *A Modern English Grammar on Historical Principles*. Part VI *Morphology*. London: George Allen & Unwin.

KJELLMER, G. 1987. "Aspects of English collocations". In Meijs, W. (ed) *Corpus Linguistics and Beyond.* Amsterdam: Rodopi: 133-140.

LIVE, A. 1973. "The *take-have* phrasal in English". *Linguistics*, 9: 31-50.

MCINNES, B.T. 2004. *Extending the Log Likelihood Measure to Improve Collocation Identification* M.S. Thesis, Department of Computer Science, University of Minnesota, Duluth.

MALMKJAER, K (ed.) 1991. *The Linguistic Encyclopaedia*. London: Routledge.

MATSUMOTO, M. 1999. "Composite predicates in Middle English". In Brinton, L.J. & M. Akimoto (eds.) *Collocational and Idiomatic Aspects of Composite Predicates in the History of English.* Amsterdam: John Benjamins: 59-96.

MEIJS, W. (ed). 1987. *Corpus Linguistics and Beyond*. Amsterdam: Rodopi.

MEL'ČUK, I. 1988. "Semantic description of lexical units in an explanatory combinatorial dictionary: Basic principles and heuristic criteria". *International Journal of Lexicography*, 1988, 1(3):165-188.

MOON, R. 1994 "The analysis of fixed expressions in text." In Coulthard, M. (ed) *Advances in Written Text Analysis.* London: Routledge: 117-135.

NATTINGER, J.R. and J.S. DECARRICO. 1992. *Lexical Phrases and Language Teaching*. Oxford: Oxford U.P.

PIVAUT, L. 1994. "Quelques aspects sémantiques d'une construction à verbe support *faire*." *Lingvisticae Investigationes*, 18/1: 49-88.

SINCLAIR, J.M. 1991. *Corpus Concordance Collocation*. Oxford: Oxford U. P.

SMADJA, F. 1993. "Retrieving Collocations from Text: Xtract." *Computational Linguistics,* 19(1): 143-177.

TROTTER, D. A. 2000. *Multilingualism in Later Medieval Britain*. Rochester: Boydell and Brewer.

Received: 8 January 2007
Revised version: 15 February 2008