An Analysis of Verb Subcategorization Frames in Three Special Language Corpora with a View towards Automatic Term Recognition

EUGENIA EUMERIDOU1, BLAISE NKWENI-AZE92 and JOHN McNAUGHT3
1Department of Information and Communication Systems, University of the Aegean, Karlovassi, Samos, Greece E-mail: evmoir@icsd.aegean.gr
2Centre for Computational Linguistics, UMIST, P.O. Box 88, Sackville Street, Manchester M60 1QD, UK E-mail: blaise@ccl.umist.ac.uk
3Department of Computation, UMIST, P.O. Box 88, Sackville Street, Manchester M60 1QD, UK E-mail: Jock@co.umist.ac.uk

Abstract. Current term recognition algorithms have centred mostly on the notion of term based on the assumption that terms are monoreferential and as such independent of context. The characteristics and behaviour of terms in real texts are however far removed from this ideal because factors such as text type or communicative situation greatly influence the linguistic realisation of a concept. Context, therefore, is important for the correct identification of terms (Dubuc and Lauriston, 1997). Based on this assumption, we have shifted our emphasis from terms towards surrounding linguistic context, namely verbs, as verbs are considered the central elements in the sentence. More specifically, we have set out to examine whether verbs and verbal syntax in particular, could help us towards the task of automatic term recognition. Our findings suggest that term occurrence varies significantly in different argument structures and different syntactic positions. Additionally, deviant grammatical structures have proved rich environments for terms. The analysis was carried out in three different specialised subcorpora in order to explore how the effectiveness of verbal syntax as a potential indicator of term occurrence can be constrained by factors such as subject matter and text type.

Key words: automatic term recognition, special languages, special language subcorpora, terms, term extraction, verb subcategorisation patterns

1. Introduction

Automatic term recognition is an area of research and an aspect of terminology which is gaining increasing prominence in our technological society, as a wide range of scientists from different disciplines – computational linguists, applied linguists, translators, interpreters and computer engineers – need to extract terminologies from texts. By terminologies we refer to the concepts that make up the conceptual framework of a domain together with their linguistic realizations,
the terms. Such a task used to be particularly laborious in the past, as term extraction was carried out wholly manually. However, the introduction of computers to language analysis has provided terminologists with huge amounts of text, the size of which necessitated automation of the search and the extraction process. As a result, a number of term extraction systems have been developed to assist terminological work.

Designers of current term recognition systems have either mapped term linguistic properties onto computational procedures, e.g. exploiting term formation patterns to extract candidate terms (Bourigault, 1992; Ananiadou, 1994; Jacquemin and Royaute, 1994; Nkwenti-Azeh, 1994; Dagan and Church, 1995; Oueslati et al., 1996) or have applied statistical techniques to measure the degree of unithood or termhood of the candidate multi-word terms (Smadja, 1991; Damerau, 1993; Haas and He, 1993; Enguehard and Pantera, 1994; Cohen, 1995). Occasionally, a combination of both techniques has been used, resulting in hybrid systems (Daille et al., 1994; Franzoi and Ananiadou, 1996; Maynard and Ananiadou, 1999). In the current state of the art, no system performs term recognition in the actual sense of the word. They simply provide researchers with lists of candidate terms which need further validation by a subject specialist. Additionally, the evaluation of such systems has proved difficult. In a review of 12 current systems of term extraction Castellvì et al. (2001) note that their evaluation is not carried out in great depth and conclude that “Broadly speaking there is neither clear nor measurable explanation of the final results. [...] it is difficult to evaluate and compare them” (p. 82). This is partly due to lack of a common test bench together with criteria to carry out such an evaluation, and in particular of a subcorpus marked-up for terms in which all terms will have been previously successfully identified, against which a given technique will be run to measure how well it has done in comparison with this “gold standard” (Kageura et al., 1998). Moreover, a system is usually trained on small and highly specialised subcorpora with regard to the topic as well as the degree of specialisation, which makes it difficult to use or test these systems in different environments. On the whole, all systems propose large lists of candidate terms which have to be manually checked for termhood. The authors end their review of current term recognition systems with a set of improvements future designers of term extraction systems should take into account to improve their performance. Among potential improvements, they mention study of the influence of the syntactic function of terminological phrases on texts and of the type of constraints terminological units present with respect to conceptual field and text type.

Additionally, recent studies (Lauriston, 1996) have stressed the importance of context for the successful recognition and disambiguation of terms, as contrary to former beliefs (Wüster, 1978) actual terminological practice has shown that terms are not independent of their linguistic context. On the contrary, they depend on context to convey their specialized meaning, as well as the form they appear in (Dubuc and Lauriston, 1997). Moving along these lines we have decided to shift focus from terms themselves to their environment and explore whether surrounding
linguistic information can help us predict term presence. As carriers of contextual information we have chosen verbs, considered by many as the central elements in the sentence and the main distributors of concepts in the sentence (Fillmore, 1968; Chafe, 1970). During our research, we have studied various ways verbs can contribute to term recognition — through their form, syntax and semantics. In this paper, we will present one aspect only, which refers to the potential contribution of verbal subcategorisation patterns towards term prediction. More specifically, we examine:

1. whether there is any frequency relationship between verb subcategorisation patterns and term occurrence, i.e. whether certain subcategorisation patterns may be more frequented by terms than others, in which case knowledge of their interrelationship could be a good indicator of environments in a subcorpus rich in terms;

2. whether there is any frequency relationship between certain syntactic positions within verb subcategorisation patterns and term occurrence, i.e. whether certain syntactic positions are more frequented by terms than others, in which case knowledge of their interrelationship could be a good indicator of environments rich in terms.

It is important to clarify that, at this point in our research, we have not aimed at the construction of an automatic term extraction or recognition system but rather are interested in establishing whether and to what extent surrounding linguistic context realised through verbs determines the presence of terminological units and thus can serve as an indicator of term presence. As our analysis is carried out on special language texts, in Section 2 we refer to studies concerning verbal syntactic behaviour in special language and point to potential ways such behaviour could help us towards term recognition. In Section 3, we present the tools and resources used in our research together with the methodology we followed to extract the results of our analysis. Section 4 presents the main argument structures encountered in each subcorpus. Sections 5, 6 and 7 discuss our findings with respect to the relation between verbal argument structure and term occurrence, syntactic position and term occurrence and deviant argument structures and term occurrence. In Section 8, we discuss our findings and how these are related to each subcorpus. Finally, Section 9 concludes our findings concerning the contribution of verbal argument structures to term recognition and how such findings are constrained by the special language factor.

2. Verbal Subcategorisation Patterns in Special Languages

In traditional terminology, terminologists were only interested in the lexicon as the sole means to express special reference. However, in recent years, researchers have pointed out that special reference is not only expressed via a specialised vocabulary which constitutes the terminology of a specific subject field but also manifests itself in syntax and semantics, thus resulting in subsystems of language, fully
operating at all linguistic levels (Lehrberger, 1986). Such language subsystems are commonly known as sublanguages or special languages and are used "(...) by a community of speakers, say, those concerned with a particular subject matter or those engaged in a specialised occupation" (Bross et al., 1972, p. 1303).

As far as syntax is concerned, researchers (Bonzi, 1990) have pointed out that special languages are often marked by preference for certain grammatical structures, while at the same time demonstrating ungrammatical structures (from a general language point of view) which constitute perfect examples of well-formed sentences for their respective special languages (Lehrberger, 1986).

As languages manifest a harmonious collaboration and interdependence among linguistic items at all linguistic levels, it would be interesting to explore whether specialised verbal behaviour entails term presence. In other words, we have set out to examine:

1. whether verbal syntactic behaviour in special languages, when deviant from or simpler than general language, would indicate environments rich in terms.
2. whether the prevalent grammatical structures in a given special language subcorpus are the richest environments for terms.

3. Methodology

To investigate the contribution of verbal subcategorisation patterns towards automatic term recognition, we have studied a wide range of verbal subcategorisation patterns in different environments based on a selection of three special language subcorpora found in the British National Corpus (BNC) (Leech, 1993).

The three sample subcorpora examined for this purpose are of comparable size and belong to different special languages both in terms of subject matter and text type. The first subcorpus, the FRT subcorpus (34,136 words), is a regulation and belongs to the special language of commerce. The second subcorpus, the CMT subcorpus (20,293 words), is a handbook and belongs to the special language of computing. Finally, the third subcorpus, the EMT subcorpus (26,467 words), is a lecture and belongs to the special language of electrical engineering.

To carry out our subcorpus analysis in the three special language subcorpora, we used Xtract (Smadja, 1991), a collocation retrieval tool which provided us with long enough contexts to study different occurrences of the same verb. The program was run in three stages:

1. During the first stage, a file was output with all the frequencies of occurrence of the verbs. Only verbs with frequency of occurrence higher than three were extracted, as verbs of a lower frequency were not considered informative or representative enough of a subcorpus' content.
2. Sentences were extracted, containing instances of these verbs.
3. For each verb, n-grams were extracted of all the words occurring within five positions both before and after the verb. This is the maximum number of positions Xtract allows. However, this last stage did not prove very useful in
our analysis, as the arguments of the verbs were frequently found at a greater distance than five preceding or following positions.

Once the sentences containing instances of each subcorpus verbs were extracted, they were manually parsed, using the syntactic patterns defined in the Oxford Advanced Learner's Dictionary of Current English (Hornby, 1974). The complete set of the fifty-one (51) OALD verb patterns and their tags is reproduced in Appendix I while the existing set of tags together with examples drawn from the three subcorpora are found in Appendix II. Tables for each argument structure were then constructed, grouped into transitive and intransitive structures according to the OALD grammatical structure classification, containing all the instances of verbs following this structure together with their argument frames.

Subsequently, the relation between verb subcategorisation patterns and term occurrence was explored along the following lines:

1. Non-use and frequencies of occurrence of different structures were considered to establish the subcorpus characteristics in terms of verbal structures;
2. A frequency analysis of term occurrence was manually carried out to establish which verb subcategorisation patterns are richest in terms in each subcorpus;
3. A frequency analysis of terms for each argument position was manually performed to determine which are the best positions for terms in each subcategorisation pattern in each subcorpus; Percentages for term occurrence in each syntactic position were calculated with respect to the total number of times such a position is filled in.
4. Deviant patterns were examined for their possible terminological value.

To decide on the termhood of nouns during the manual annotation in our research, we consulted expert opinion as well as a variety of dictionaries, e.g. A Dictionary of Finance (Butler and Isaacs, 1993), Dictionary of Commercial, Financial and Legal Terms (Herbst, 1966), Dictionary of Electrical Engineering (Jackson and Feinberg, 1981), A Dictionary of Law (Curzon, 1982), Elsevier's Dictionary of Personal and Office Computing (Vollnhals, 1984), and The 3-D Visual Dictionary of Computing (Graham, 1995). At this point, it should be noted that our approach to term definition is the pragmatic one (Pearson, 1998). That is, a term is any linguistic item which has a specialised usage in our corpus, regardless of whether it belongs to the vocabulary of the given special language or of different special languages, or it is a general language word which has acquired term status through a restriction of its meaning. As a result, in the third subcorpus we analysed, we have included as terms items which do not belong strictly to the special language of electrical engineering but to the special language of maths and statistics as well. Additionally, in the second special language subcorpus, we note a wide number of words which are general language words e.g. line, character, page, format but which have been counted as terms, as they have acquired terminological meaning through a restriction of their meaning.
4. The Distribution of Verbal Subcategorisation Patterns in the Three Special Language Subcorpora

In this section, we examine the distribution of verbal subcategorisation patterns in the three special language subcorpora and establish the prevalent subcategorisation patterns in each of them. Table I displays the frequencies of intransitive structures in the different subcorpora together with their respective percentages.

According to Table I, the main intransitive structures are all common in the three special language subcorpora. Among the three commonest structures, we have $S + vi$, $S + vi + \text{preposition} + \text{noun}$ and $S + vi + \text{adjective/noun}$. However, the frequency of the above grammatical structures varies widely in the different subcorpora. For example, structure $S + vi$ occupies 75% of the grammatical structures in the CMT subcorpus and only 35% in the FRT subcorpus. It is also worth mentioning the simplicity of structures that the CMT subcorpus presents. Thus, structure $S + vi$ occupies the highest percentage while the remaining space is occupied by five additional structures. On the contrary, the other two subcorpora have the major distribution of occurrences under three structures and a wide number of smaller structures. The simplicity of grammatical structures that the CMT subcorpus presents is useful for term recognition as it concentrates our search for terms mainly on the syntactic position of one structure. Table II displays the frequencies of the main transitive structures in the different subcorpora, together with their respective percentages.

Table II shows the main transitive structures in the three special language subcorpora. The largest number of structures is found in the FRT subcorpus, while frequencies are evenly distributed in a much wider number of structures than in the other two subcorpora. The EMT subcorpus presents a smaller number of grammatical structures while the majority of occurrences is split between two grammatical structures. The CMT subcorpus has only one prevalent structure, namely $S + vt + \text{noun}$. The rest appear in insignificant proportions. We note from the table that patterns $S + vt + \text{noun}$, $S + vt + \text{noun} + \text{adverbial phrase}$, $S + vt + \text{that-clause}$ and $S + vt + \text{noun} + \text{preposition} + \text{noun}$ constitute the predominant structures in the three subcorpora when considered as a whole. However, each corpus has its own characteristic structures.2 The FRT subcorpus has the structures $S + vt + \text{noun} + \text{preposition} + \text{noun}$ and $S + vt + \text{noun} + \text{adverbial phrase}$. It is also worth mentioning the presence of the structure $S + vt + \text{noun} + \text{to} + \text{noun}$, which, although it presents a low rate of occurrence, is missing from the other corpora. In the CMT subcorpus, the main structures are $S + vt + \text{noun} + \text{adverbial phrase}$ and $S + vt + \text{noun} + \text{adverbial particle}$. Finally, the EMT subcorpus presents as its main structure the structure $S + vt + \text{noun} + \text{adverbial phrase}$. Additionally, all corpora present at a high frequency rate the structure $S + vt + \text{that-clause}$. This structure is characteristic of the text type of the three subcorpora as it is a structure closely connected to communication verbs and such verbs are important in all subcorpora, as the first one is a regulation, the second is a handbook and the third is a lecture.
Table I. Intransitive verb subcategorisation patterns in the three subcorpora

<table>
<thead>
<tr>
<th>FRT Construction</th>
<th>Freq</th>
<th>Total</th>
<th>CMT Construction</th>
<th>Freq</th>
<th>Total</th>
<th>EMT Construction</th>
<th>Freq</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>S + vi + preposition + noun/pronoun</td>
<td>302</td>
<td>51%</td>
<td>S + vi + (adverbal phrase)</td>
<td>49</td>
<td>74%</td>
<td>S + vi + (adverbal phrase)</td>
<td>119</td>
<td>43%</td>
</tr>
<tr>
<td>S + vi + (adverbal phrase)</td>
<td>207</td>
<td>35%</td>
<td>S + be + complement/adjunct</td>
<td>5</td>
<td>7.5%</td>
<td>S + vi + preposition + noun/pronoun</td>
<td>109</td>
<td>39%</td>
</tr>
<tr>
<td>S + vi + adjective/noun/pronoun</td>
<td>60</td>
<td>10%</td>
<td>S + vi + preposition + noun/pronoun</td>
<td>5</td>
<td>7.5%</td>
<td>S + vi + adjective/noun/pronoun</td>
<td>31</td>
<td>11%</td>
</tr>
<tr>
<td>Other</td>
<td>23</td>
<td>4%</td>
<td>S + vi + adjective/noun/pronoun</td>
<td>3</td>
<td>4.5%</td>
<td>S + be + complement/adjunct</td>
<td>10</td>
<td>4%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>S + vi + to-infinite</td>
<td>2</td>
<td>3%</td>
<td>Other</td>
<td>9</td>
<td>3%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>S + finite of BE + to-infinite</td>
<td>2</td>
<td>3%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the following section, only the argument structures occurring more than ten times will be examined in relation to termhood.

5. The Relation of Verbal Subcategorisation Patterns and Term Occurrence in the Three Special Language Subcorpora

In this section, we examine the distribution of terms in the argument structures with frequency higher than ten occurrences in order to explore:

1. whether there is any frequency relationship between verbal subcategorisation patterns and term occurrence;
2. whether the prevalent grammatical structures in a given special language subcorpus are the richest environments for terms.

Tables III and IV show term distribution in the three special language subcorpora. The percentage figure relates to the occurrence of a term in any of the argument positions.

According to Table III, term distribution varies across different intransitive argument structures within the same subcorpus. For instance, structure $S + vi +$
Table II. Transitive verb subcategorisation patterns in the three subcorpora

<table>
<thead>
<tr>
<th>Construction</th>
<th>FRT</th>
<th>CMT</th>
<th>EMT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Freq</td>
<td>Total</td>
<td>Freq</td>
</tr>
<tr>
<td>S + vt + noun/ pronoun</td>
<td>1407</td>
<td>49%</td>
<td>S + vt + noun/ pronoun</td>
</tr>
<tr>
<td>S + vt + noun + preposition + noun</td>
<td>267</td>
<td>10%</td>
<td>S + vt + noun + adverbial phrase</td>
</tr>
<tr>
<td>S + vt + that-clause</td>
<td>266</td>
<td>10%</td>
<td>S + vt + noun + adverbial particle</td>
</tr>
<tr>
<td>S + vt + noun + adverbial phrase</td>
<td>243</td>
<td>9%</td>
<td>S + vt + that-clause</td>
</tr>
<tr>
<td>S + vt + noun + to-infinite</td>
<td>192</td>
<td>7%</td>
<td>S + vt + to infinitive</td>
</tr>
<tr>
<td>S + vt + to-infinite</td>
<td>106</td>
<td>4%</td>
<td>S + HAVE/ OUGHT + to-infinite</td>
</tr>
<tr>
<td>S + vt + noun + to + noun</td>
<td>80</td>
<td>3%</td>
<td>S + vt + noun + preposition + noun</td>
</tr>
<tr>
<td>S + vt + noun + adverbial particle</td>
<td>68</td>
<td>2%</td>
<td>S + vt + noun + to-infinite</td>
</tr>
<tr>
<td>S + vt + noun/ pronoun/ gerund + adjective</td>
<td>51</td>
<td>2%</td>
<td>Other</td>
</tr>
<tr>
<td>Other</td>
<td>114</td>
<td>4%</td>
<td></td>
</tr>
</tbody>
</table>

preposition + noun presents a term rate of 78% in the EMT subcorpus, whereas structure S + be + complement/adjunct presents only a 30% term rate. Additionally, term distribution varies within the same argument structure across different subcorpora. Thus, the argument structure S + vi presents a 86% rate for terms in the FRT subcorpus, a 67% rate for terms in the EMT subcorpus and only a 39% rate for terms in the CMT subcorpus.
Table III. Term distribution in the main intransitive verb subcategorisation patterns

<table>
<thead>
<tr>
<th></th>
<th>FRT Construction</th>
<th>Total%</th>
<th>CMT Construction</th>
<th>Total%</th>
<th>EMT Construction</th>
<th>Total%</th>
</tr>
</thead>
<tbody>
<tr>
<td>S + vi + preposition + noun/pronoun</td>
<td>62%</td>
<td></td>
<td>S + vi</td>
<td>39%</td>
<td>S + vi + preposition + noun/pronoun</td>
<td>78%</td>
</tr>
<tr>
<td>S + vi + (adverbial phrase)</td>
<td>86%</td>
<td></td>
<td></td>
<td></td>
<td>S + vi + (adverbial phrase)</td>
<td>67%</td>
</tr>
<tr>
<td>S + vi + adjective/noun</td>
<td>87%</td>
<td></td>
<td></td>
<td></td>
<td>S + vi + adjective/noun</td>
<td>64%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>S + be + complement/adjunct</td>
<td></td>
<td></td>
<td>30%</td>
</tr>
</tbody>
</table>

The same observations are derived from Table IV. Term distribution widely varies across different argument structures within the same special language subcorpus and within the same structure across different special language subcorpora. For instance, in the CMT subcorpus, structure $S + vt + noun + adverbial particle$ presents a term rate of 90%, whereas structure $S + vt + that-clause$ presents a 0% term rate. Additionally, in the CMT subcorpus, structure $S + vt + noun + preposition + noun$ presents a rate of 45% whereas the same structure in the EMT subcorpus presents a term rate of 86%. Finally, we note that the main structures in the three subcorpora are among the richest environments in terms. Thus, in the FRT subcorpus, structures $S + vt + noun + preposition + noun$ and $S + vt + noun + adverbial phrase$ present a term rate of 79%, in the CMT subcorpus structures $S + vt + noun + adverbial phrase$ and $S + vt + noun + particle + noun$ present a term rate of 90% and 91% respectively and finally in the EMT subcorpus structure $S + vt + noun + adverbial phrase$ presents a term rate of 85%. It is also worth mentioning that structures $S + vt + noun + to + noun$ in the FRT and $S + vt + noun/pronoun/gerund + adjective$ in the EMT subcorpus present surprisingly high term rates.

The above findings suggest that term distribution varies among different argument structures and that knowledge of which structures are extremely rich or poor in terms could help guide us to the best search environments for an automatic term recognition algorithm.

6. The Relation of Syntactic Positions to Term Occurrence in the Three Special Language Subcorpora

Now that the prevalent term formation patterns in terms of frequency and term occurrence have been established in each subcorpus, we examine the relation of syntactic positions and term occurrence within these argument structures in the
Table IV. Term distribution in the main transitive verb subcategorisation patterns

<table>
<thead>
<tr>
<th>Construction</th>
<th>CMT</th>
<th>EMT</th>
</tr>
</thead>
<tbody>
<tr>
<td>S + vt + noun</td>
<td>72%</td>
<td>73%</td>
</tr>
<tr>
<td>pronoun</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S + vt + that-clause</td>
<td>47%</td>
<td>91%</td>
</tr>
<tr>
<td>S + vt + noun + adverbial phrase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S + vt + noun + preposition + noun</td>
<td>79%</td>
<td>90%</td>
</tr>
<tr>
<td>S + vt + that-clause</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S + vt + noun + adverbial phrase</td>
<td>79%</td>
<td>90%</td>
</tr>
<tr>
<td>S + vt + noun + to-infinite</td>
<td>58%</td>
<td>6%</td>
</tr>
<tr>
<td>S + vt + to-infinite</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S + vt + noun + to + noun</td>
<td>100%</td>
<td>9%</td>
</tr>
<tr>
<td>S + HAVE/OUGHT + to-infinite</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S + vt + noun + adverbial particle</td>
<td>63%</td>
<td>45%</td>
</tr>
<tr>
<td>S + vt + noun + preposition + noun</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S + vt + to-infinite</td>
<td>53%</td>
<td></td>
</tr>
<tr>
<td>S + vt + noun/ pronoun/gerund + adjective</td>
<td>43%</td>
<td></td>
</tr>
</tbody>
</table>

three special language subcorpora. Table V shows term distribution in the various syntactic positions within each prevalent intransitive argument structure in the three subcorpora.

Table V shows that knowledge of argument position is important in determining the term status of a given word. For example, in the FRT subcorpus, the subject position in argument structure S + vi + adverbial phrase has a likelihood of 69% to be filled by a term, whereas the adverbial head position has a likelihood of 43% to be occupied by a term. Additionally, we note that the importance of a syntactic position as a probe for terms largely depends on the special language factor. Thus, the subject position in the FRT subcorpus is a fairly good position for terms, whereas the same position is the poorest candidate for termhood in the CMT subcorpus. Table VI below shows the best positions within the prevalent transitive argument structures for term recognition in the three special language subcorpora.
**Table V.** Distribution of terms in the argument positions of intransitive structures across the subcorpora

<table>
<thead>
<tr>
<th>Structure</th>
<th>Argument position</th>
<th>% Terms in Subcorpora</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>FRT</td>
</tr>
<tr>
<td>S + vi + (adverbial phrase)</td>
<td>Subject</td>
<td>111 (69%)</td>
</tr>
<tr>
<td></td>
<td>Adverbial phrase</td>
<td>60 (43%)</td>
</tr>
<tr>
<td>S + vi + adjective/noun</td>
<td>Subject</td>
<td>41 (78%)</td>
</tr>
<tr>
<td></td>
<td>Predicative complement</td>
<td>42 (70%)</td>
</tr>
<tr>
<td>S + vi + preposition + noun</td>
<td>Subject</td>
<td>212 (76%)</td>
</tr>
<tr>
<td></td>
<td>Prepositional phrase</td>
<td>195 (65%)</td>
</tr>
</tbody>
</table>

**Table VI.** Distribution of terms in the argument positions of transitive structures across the subcorpora

<table>
<thead>
<tr>
<th>Structure</th>
<th>Argument position</th>
<th>% Terms in Subcorpora</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>FRT</td>
</tr>
<tr>
<td>S + vt + noun/pronoun</td>
<td>Subject</td>
<td>588 (74%)</td>
</tr>
<tr>
<td></td>
<td>Object</td>
<td>850 (63%)</td>
</tr>
<tr>
<td>S + vt + to-infinite</td>
<td>Subject</td>
<td>56 (53%)</td>
</tr>
<tr>
<td>S + vt + that-clause</td>
<td>Subject</td>
<td>128 (51%)</td>
</tr>
<tr>
<td>S + vt + noun/pronoun +</td>
<td>Subject</td>
<td>40 (72%)</td>
</tr>
<tr>
<td>preposition + noun/pronoun</td>
<td>Direct object</td>
<td>144 (60%)</td>
</tr>
<tr>
<td></td>
<td>Prepositional object</td>
<td>191 (71%)</td>
</tr>
<tr>
<td>S + vt + noun/pronoun +</td>
<td>Subject</td>
<td>37 (57%)</td>
</tr>
<tr>
<td>adverbial phrase</td>
<td>Direct object</td>
<td>159 (69%)</td>
</tr>
<tr>
<td></td>
<td>Adverbial head</td>
<td>179 (69%)</td>
</tr>
<tr>
<td>S + vt + adverbial particle +</td>
<td>Subject</td>
<td>27 (55%)</td>
</tr>
<tr>
<td>noun/pronoun</td>
<td>Direct object</td>
<td>43 (63%)</td>
</tr>
</tbody>
</table>

It is evident from the table that term distribution varies depending on the argument position of a given structure. In the structure $S + vt + noun$, which is the most numerous one, the object position is a fairly good environment for terms in all three special language subcorpora. Structures $S + vt + to\text{ }infinitive$ and $S + vt + that\text{ }clause$ are structures poor in terms, except for the subject position in the FRT subcorpus. Structures $S + vt + noun + preposition + noun$ and $S + vt + noun + adverbial\text{ }phrase$ are the best argument structures for term occurrence as they
have terms in all three argument positions. Reliable candidates are words falling into the direct object, prepositional object and adverbial head position. Finally, for structure S + vt + noun + adverbial particle, the best candidates for termhood are words falling into the object position. As for the special language factor, we note that, depending on the special language subcorpus different positions acquire different importance for term recognition. The most striking example is the subject position in the FRT subcorpus, which is the best position for termhood, while in the other two subcorpora it is the poorest.

The evidence from Tables V and VI suggests that knowledge of the argument position in a certain grammatical structure is important for predicting the term status of a given noun.

7. Deviant Syntactic Structures and Term Occurrence

Deviant grammatical structure is an additional indicator of special language usage. Our last aim was to investigate whether such environments would be rich in term occurrence. The three subcorpora we have examined, taken from the BNC, are all labelled as books according to the BNC text type classification. As a result, we have few omissions and ungrammatical constructions. A thorough subcorpus analysis, however, has yielded examples of deviant usage which will be presented below to show their relation to term occurrence.

7.1. Deviant Grammatical Structures in the FRT Subcorpus

Starting with the first subcorpus, the FRT subcorpus, we note that the majority of grammatical structures are marked by their regularity. The limited number of unusual grammatical constructions is presented below (items underlined are terms).

Sell back.
He has sold cold.
To close out a position.
The contract falls to be performed.

According to the frequency analysis of English vocabulary and grammar, based on the LOB subcorpus (Johansson, 1989), fall does not occur with an infinitive in 279 occurrences, close can be followed by a preposition but not by out and sell is rarely followed by an adverb or adjective at all. The above examples are instances of special language constructions and they are rich environments for terms as the underlined words show. It is often the case in such constructions that the verb itself is used as a technical expression. For example, the verb close is a general language verb, used as term. Thus, to close out a position means to close an open position, usually on a futures market by buying to cover a short sale or by selling a long purchase.
7.2. Deviant Grammatical Structures in the CMT Subcorpus

The CMT subcorpus presents us with the following set of examples in which a verb is followed by a verb.

- **Select** Copy.
  - to **select** PRINT REPAGINATE confirm page breaks.
  - to **select** FORMAT DECISION PAGE NUMBERS.
  - **tap** ENTER.
  - **tap** Overtypes.
  - **hold down** Shift.
  - **tap enter** to end the table.
  - **tap enter** to finish the table.
  - **tap enter** to force a page break.

In the above examples, both verbs in the sentence are terms. The first verb is an imperative and functions as the main verb in the sentence; the second verb, meanwhile, is used as a noun although it retains its verbal form. This is evident in cases where the nominal form is different from the verbal, e.g. *enter* (v) – *entry* (n). We can also rephrase the commands to illustrate this functional difference, e.g. "**Tap the key called ENTER; Select the command named Copy**".

7.3. Deviant Grammatical Structures in the EMT Subcorpus

The third subcorpus, the EMT subcorpus, appears to be regular in grammatical structures. The only example of unusual grammatical structure is displayed by the verb *decay* in the following examples:

- The second term decays as 1 r.
- The potential decays as 1 r.
- The integrand decays as 1 r.

In the above sentences, the verb *decay* is followed by the preposition *as* and subcategorises for a noun expressing an abstract entity in subject position, whereas it normally occurs on its own or it is followed by an adverbial and subcategorises for a word meaning food, a living part of an organism or anything that can physically decay. In the electrical engineering special language, however, *decay* is a verb term and means *decline*. As the examples show, it is an environment rich in terms. The above examples suggest that deviant grammatical structures are indicators of terminological usage assigning term status either to the verb itself or to its surrounding arguments.

8. Discussion

In this paper, we set out to examine whether there is any correlation between argument structure and term occurrence in order to improve the performance of current term recognition systems. To achieve our goal, we firstly examined term occurrence
with respect to argument structure and then with respect to syntactic position. Additionally, we examined whether the special language factor could enhance the effectiveness of verbal subcategorisation patterns as potential indicators of termhood. More specifically, we examined whether prevalent or characteristic argument structures in different special language subcorpora were richer environments for terms than others and whether the existence of deviant grammatical structures could indicate term occurrence.

Our findings showed that indeed term occurrence varies widely depending on the argument structure and on the syntactic position. Additionally, the special language factor plays an important role as it determines which are the best syntactic environments for term occurrence. To start with, from the initial set of 51 argument structures described in OALD dictionary, we encounter on average only 15 in each subcorpus, while the most significant ones in terms of frequency are on average 5. A distribution analysis of the argument structures in each subcorpus shows that the most frequent argument structures are common in all subcorpora. However, they rank differently and this shows their importance in each subcorpus. Subcorpus analysis has shown that the prevalent structures are good environments for terms. For instance, structures $S + vi$ and $S + vi + preposition + noun$ are among the prevalent and richest in terms intransitive structures both for the FRT and the EMT subcorpus. In the FRT subcorpus, this is due to the high frequency of intransitive occurrences of the verbs trade and deal which are terms themselves in the subcorpus and largely surrounded by terms. In the EMT subcorpus, it is due to the fact that intransitive verbs are used to describe processes, states and changes which involve substances, events and phenomena which in the EMT subcorpus are terms. Moreover, reliable results in terms of term occurrence may also come from structures which have an average frequency in the subcorpus, but are missing from the other subcorpora. Such argument structures are often characteristic of a given special language. For instance, the FRT subcorpus, which falls into the commercial special language abounds in commercial terms which usually follow verbs of giving. The structure most closely related to these verbs is the structure $S + vt + noun + to + noun$. It is worth noticing that this structure is missing from other subcorpora and has 100% term occurrence, e.g. He must pay the contract price to a selling member, a fiduciary owes duties to the beneficiary. In the EMT subcorpus, an argument structure indicative of its subject matter is the structure $S + vt + noun + gerund/adjunctive/noun$. This structure presents a surprisingly high rate of terms compared to the rest of the subcorpora and this is due to the fact that the EMT subcorpus is rich in single word adjective terms, thus providing an additional position for term occurrence, e.g. We shall keep the voltage constant, a rectangular coil is located parallel etc. Finally, in the CMT subcorpus, we have the structure $S + vt + particle + noun$ which is important as it abounds in phrasal verbs, e.g. hold down, set up, type in which have a specialized usage in the CMT subcorpus and as such they are good environments for terms, e.g. to set up the format, you hold down shift, you hold down control, you switch on the toggle.
The particular subject matter a subcorpus belongs to influences the way terms are distributed in the argument structure as well. For instance, the subject position in transitive structures is a good position for terms only in the FRT subcorpus. This is due to the fact that the subcorpus abounds in terms referring to commercial institutions, regulatory bodies and instruments and these words carrying the feature of potency frequently occur in subject position. The opposite is the case for the other subcorpora as the subject position is mostly empty or occupied by a pronoun. In the CMT subcorpus, the best positions are the object position and the adverbial head position as terms in this subcorpus refer to entities perceived, touched or created as well as entities expressing location. In the EMT subcorpus, the best position for candidate terms is the subject position for intransitive verbs as there mostly fall words referring to substances, material and phenomena which are terms. Additionally, the adverbial head position is a good position for terms as also words denoting location are usually terms. Finally, we examined the importance of deviant structures for term recognition in the three special language subcorpora and we found that unusual grammatical structures are rich environments in terms.

An additional filter is provided by text type. The only argument structure which can be related to text type in the three subcorpora is the structure \( S + vt + that \) clause as we have already mentioned in section 4. However, this structure is a poor environment for terms, in all three subcorpora, rendering the factor of text type an irrelevant factor in terms of term recognition via verbal argument structures.

9. Conclusions

In the above sections, we set out to examine (a) whether knowledge of the grammatical constructions and argument position can help us predict the term status of a given noun and (b) whether knowledge of the prevalent grammatical constructions or deviant structures of a given special language can help us in term recognition. The above findings suggest that term distribution can vary widely across different structures or argument positions. Such knowledge could be important for automatic term recognition both as an indicator of term presence and as an indicator of word presence. They also show that the special language factor is crucial for the effectiveness of our approach. To start with, it determines whether argument structures can play a role as indicators of term occurrence or not. For instance, corpora, such as the CMT subcorpus, which have a very low number of argument structures, but where there is a high correlation between argument structures and term occurrence, would prove ideal for our approach. On the other hand, in case of a low correlation between prevalent verbal argument structures and term occurrence, such a corpus would be prohibitive as having few argument structures it would provide us with few alternatives. Additionally, the special language a subcorpus belongs to determines our search environments for terms, as each special language corpus has its own preferred grammatical structures and ways of term distribution in these structures.
To conclude, our results, though promising, are of an indicative nature only. The most significant aspect of our research is that we have shown verbs add a new dimension in our efforts to extract terminological units from texts. The findings presented in the previous sections form the basis of further research in this direction. In particular, we can
(a) analyse other, larger subcorpora in the same domain to arrive at domain-level universals, i.e. study of verbal contribution to term recognition in subcorpora belonging to one special language, e.g. commercial, but to different text types to see whether we can make generalisations about the behaviour of certain verbs in a given domain irrespective of text-type;
(b) analyse other, larger subcorpora of the same text type to arrive at text-type universals, i.e. study of verbal contribution to term recognition in subcorpora belonging to the same text type but to different special languages;
(c) incorporate the results in an automatic term recognition system, by examining the most productive verbal argument structures for term recognition in each special language subcorpus.

Appendix I

Appendix I presents the complete set of the verb patterns used in verb argument structure analysis as identified in the OALD (Hornby, 1974, p. xvi). The abbreviations used are: S = Subject, vi = intransitive verb, vt = transitive verb, DO = Direct Object, IO = Indirect Object, anom fin = anomalous finite.

(1) [VP1] S + BE + complement/adjunct
(2) [VP2A] S + vi
(3) [VP2B] S + vi + (for) adverbial adjunct
(4) [VP2C] S + vi + adverbial adjunct etc.
(5) [VP2D] S + vi + adjective/noun/pronoun
(6) [VP2E] S + vi + present participle/participial phrase
(7) [VP3A] S + vi + preposition + noun/pronoun/gerund
(8) [VP3B] S + vi + (preposition + it) + clause, etc.
(9) [VP4A] S + vi + to-infinite (phrase)
(10) [VP4B] S + vi + to-infinite (phrase)
(11) [VP4C] S + vi + to-infinite (phrase)
(12) [VP4D] S + SEEM/APPEAR, etc. + (to be) + adjective/noun
(13) [VP4E] S + SEEM/APPEAR/HAPPEN/CHANCE + to-infinite (phrase)
(14) [VP4F] S + finite of BE + to-infinite (phrase)
(15) [VP5] S + anom fin + infinitive (phrase)
(16) [VP6A] S + vt + noun/pronoun
(17) [VP6B] S + vt + noun/pronoun
(18) [VP6C] S + vt + gerund, etc.
(19) [VP6D] S + vt + gerund, etc.
(20) [VP6E] S + NEED/WANT, etc. + gerund, etc. (passive)
(21) [VP7A] S + vt + (not) + to-infinite, etc.
(22) [VP7B] S + HAVE/ought, etc. + (not) + to-infinite
(23) [VP8] S + vt + interrogative pronoun/adverb + to-infinite
(24) [VP9] S + vt + that-clause
(25) [VP10] S + vt + dependent clause/question
(26) [VP11] S + vt + noun/pronoun + that-clause
(27) [VP12A] S + vt + noun/pronoun (IO) + noun/pronoun (DO)
(28) [VP12B] S + vt + noun/pronoun (IO) + noun/pronoun (DO)
(29) [VP12C] S + vt + noun/pronoun + noun/pronoun
(30) [VP13A] S + vt + noun/pronoun (DO) + to + noun/pronoun (phrase)
(31) [VP13B] S + vt + noun/pronoun (DO) + for + noun/pronoun (phrase)
(32) [VP14] S + vt + noun/pronoun (DO) + preposition + noun/pronoun (phrase)
(33) [VP15A] S + vt + noun/pronoun (DO) + adverbial phrase
(34) [VP15B] S + vt + noun/pronoun (DO) + adverbial particle / S + vt + adverbial particle + noun/pronoun (DO)
(35) [VP16A] S + vt + noun/pronoun (DO) + to-infinite (phrase)
(36) [VP16B] S + vt + noun/pronoun (DO) + as/like + noun (phrase) or clause
(37) [VP17] S + vt + noun/pronoun (DO) + (not) + to-infinite
(38) [VP18A] S + vt + noun/pronoun + to-infinite (phrase)
(39) [VP18B] S + vt + noun/pronoun + to-infinite (phrase)
(40) [VP18C] S + HAVE + noun/pronoun + infinitive (phrase)
(41) [VP19A] S + vt + noun/pronoun + present participle (phrase)
(42) [VP19B] S + vt + noun/pronoun + present participle (phrase)
(43) [VP19C] S + vt + noun/pronoun/possessive + ing form of the verb
(44) [VP20] S + vt + noun/pronoun + interrogative + to-infinite (phrase)
(45) [VP21] S + vt + noun/pronoun + dependent clause/question
(46) [VP22] S + vt + noun/pronoun/gerund (DO) + adjective
(47) [VP23] S + vt + noun/pronoun (DO) + noun (phrase)
(48) [VP24A] S + vt + noun/pronoun (DO) + past participle (phrase)
(49) [VP24B] S + HAVE + noun/pronoun (DO) + past participle (phrase)
(50) [VP24C] S + HAVE/GET + noun/pronoun (DO) + past participle (phrase)
(51) [VP25] S + vt + noun/pronoun (DO) + (to be) + adjective/noun
Appendix II

Appendix II presents the twenty-three (23) main subcategorisation patterns encountered in the three special language subcorpora together with examples drawn from the three subcorpora.

[VP1] S + BE + complement/adjunct

This pattern is for the verb be: The complement may be a noun, a pronoun, an adjective, an adjective phrase (e.g. a prepositional phrase). Additionally it may be an adverbial adjunct or infinitive phrase e.g. your deletion was a mistake (CMT subcorpus), the search was successful (CMT subcorpus), the material being a perfect diameter (EMT subcorpus).

[VP2A] S + vi

This pattern consists of a subject followed by an intransitive verb, followed by an optional adverbial phrase, e.g. A fiduciary relationship arises (FRT subcorpus), the broker deals on a discretionary basis (FRT subcorpus), two new symbols will appear in typed text (CMT subcorpus), the letters Column Select appear in the Status line (CMT subcorpus), electron flows between parallel plates (EMT subcorpus), the magnetic field varies (EMT subcorpus).

[VP2D] S + vi + adjective/noun/pronoun

Verbs in this pattern are followed by an adjective, a noun, or in the case of a reflexive verb a pronoun. Many inchoative verbs and verbs of senses fall under this pattern, e.g. contracts remain open (FRT subcorpus), LCH becomes counterparty (FRT subcorpus), the formulae look quite reasonable (EMT subcorpus), the electric field remained unchanged (EMT subcorpus), egn becomes an equality formula (EMT subcorpus). It should be noted at this point, that although a construction such as S + vi+ noun may seem unexpected, it can occur when the noun is used in a predicative sense, e.g. Peter is a student. Besides, the above guidelines have been consistently applied across the data. As a result individual differences in the interpretation of specific examples does not affect the overall conclusions.

[VP3A] S + vi + preposition + noun/pronoun/gerund

Verbs of this pattern are followed by a preposition plus its object. Within this pattern, we also include sentences of the form S +vi + as + noun/pronoun. The object can be a noun, a pronoun, a gerund, a phrase or a clause. Verb and preposition function as a single unit, e.g. the firm enters into speculative trades (FRT subcorpus), restrictions operate as a matter of fiduciary law (FRT subcorpus), the
highlight had been depending on the key tapped (CMT subcorpus), electromagnetic phenomena depend on the true electric current (EMT subcorpus), the above equation reduces to formula (EMT subcorpus).

[VP6A] S + vt + noun/pronoun

This is a transitive verbal pattern where the verb takes a noun/pronoun as a transitive direct object, e.g. the lawyer customer cases indicate the likely attitude of the courts (FRT subcorpus), the fiduciary owes duties (FRT subcorpus), the backspace key will not delete text (CMT subcorpus), a DBMS can retrieve this information (CMT subcorpus), the ring carries a formula (EMT subcorpus), the potential function satisfies Laplace equation (EMT subcorpus).

[VP7A] S + vt + (not) + to-infinite

In this pattern the object of the verb is to-infinite, e.g. any market claims to be an exchange (FRT subcorpus), the SFA rules seek to prevent abuse (FRT subcorpus), you want to save files (CMT subcorpus), you need to use several discs (CMT subcorpus), Derive expressions needed to hold constant (EMT subcorpus), the flux cut by each section needs to be summed algebraically (EMT subcorpus).

[VP9] S + vt + that-clause

The object of the verb is a that-clause. That is often omitted, except after more formal verbs like decide, believe etc., e.g. the traders on the floor register that a change has taken place (FRT subcorpus), a client claims that he was a private investor (FRT subcorpus), some say they are the shape of things (CMT subcorpus), you notice that something has gone wrong (CMT subcorpus), we believe you have come across the Cauchy-Rieman relationships (EMT subcorpus), we assume that the charge is uniformly distributed (EMT subcorpus).

[VP10] S + vt + dependent clause/question

In this pattern the object of the verb is a dependent clause or question. The clause is introduced by a relative pronoun or adverb what, whether/if, e.g. a sufficient range of participants and rules and procedures lay down how trades are executed (FRT subcorpus), it will determine whether churning has occurred (FRT subcorpus), to check how much is a paragraph indent (CMT subcorpus), see how well your office does (CMT subcorpus).

[VP11] S + vt + noun/pronoun + that-clause

The verb is followed by a noun or pronoun and a that-clause, e.g. the customer has not informed the firm that he wants to be treated as an indirect customer (FRT
subcorpus), the firm notifies the exchange or clearing house that the firm is (..) (FRT subcorpus).

[VP12A] S + vt + noun/pronoun (IO) + noun/pronoun (DO)

The verb is followed by an indirect and a direct object. The indirect object may be equivalent to a prepositional adjunct with to or for, e.g. The firm has given the customer a clear written warning (FRT subcorpus), it owes the putative customer fiduciary duties (FRT subcorpus), the integral of the current density gives the current the above equation (EMT subcorpus).

[VP13A] S + vt + noun/pronoun (DO) + to + noun/pronoun (phrase)

In this pattern the verb is followed by a direct object, a preposition to and the prepositional object, e.g. They owe conflicting duties to different customers, loans or credits may be granted by a firm to the customer (FRT subcorpus), the changing magnetic field can give rise to an electric field (EMT subcorpus).

[VP14] S + vt + noun/pronoun (DO) + preposition + noun/pronoun (phrase)

In this pattern the verb is followed by a direct object and a preposition and its object, e.g. The firm arranges a contingent liability transaction for a customer (FRT subcorpus), the firm arranges a contingent liability transaction with a private customer (FRT subcorpus), to reformat the text into columns (CMT subcorpus), the whole space is filled with a material of conductivity formula (EMT subcorpus), we replace the mathematical curve by thin wire (EMT subcorpus).

[VP15A] S + vt + noun/pronoun (DO) + adverbial phrase

In this pattern the direct object is followed by an adverbial phrase of place, duration, distance, etc., e.g. The broker dealer department sells stock off its own book (FRT subcorpus), Front running brings the exchange into disrepute (FRT subcorpus), insert the program disk into drive A (CMT subcorpus), you could use Ctrl for standard letters (CMT subcorpus), All the charge is concentrated at the origin of the coordinate system (EMT subcorpus), a charge of opposite sign is placed in the mirror position (EMT subcorpus).

[VP15B] S + vt + noun/pronoun (DO) + adverbial particle

S + vt + adverbial particle + noun/pronoun (DO)

In this pattern the direct object may be preceded or followed by an adverbial particle, e.g. he is giving up protections (FRT subcorpus), Parties set out the parameters of their own rights (FRT subcorpus), you do not type in the fullname (CMT subcorpus), you hold down the shift key (CMT subcorpus), you have come
across the Cauchy-Riemann relationships (EMT subcorpus), surface currents are set up (EMT subcorpus).

[VP16A] S + vt + noun/pronoun (DO) + to-infinitive (phrase)

In this pattern the verb is followed by the direct object and an infinitive of purpose or result, e.g. Clearing services enable a recognised investment exchange to make arrangements with it (FRT subcorpus), Public law enables the rules to impact on the respective obligations and rights (FRT subcorpus), move the page number to keep the two in alignment (CMT subcorpus), tap enter to force a page break (CMT subcorpus), we take the trouble to make a permanent magnet (EMT subcorpus).

[VP16B] S + vt + noun/pronoun (DO) + as/like + noun (phrase) or clause

The direct object is followed by an adjunct with as or like, or a clause introduced by if or as though, e.g. The obligations of LCH perform the contract as principal (FRT subcorpus), the broker enters into a contract as agent (FRT subcorpus), insert it as a separate clause (CMT subcorpus), we can regard the electric field as the agent (EMT subcorpus), the potential due to two point charges may be written as formula (EMT subcorpus).

[VP17] S + vt + noun/pronoun (DO) + (not) + to-infinitive

In this pattern the verb is followed by a noun/pronoun and a to-infinitive or infinitive phrase, e.g. These regulations permit calls to be made (FRT subcorpus), the firm asks the customer to waive (...) (FRT subcorpus), many commands require the Ctrl Shift or Alt keys to be held down (CMT subcorpus), the rest require one key to be held down (CMT subcorpus), we shall take the flux density to be equal (EMT subcorpus), the flux density over the wires is assumed to be infinitely thin (EMT subcorpus).

[VP18B] S + vt + noun/pronoun + infinitive (phrase)

This pattern is for verbs which are followed by bare infinitive but they are not verbs of physical perception, e.g. make a machine respond (CMT subcorpus).

[VP19B] S + vt + noun/pronoun + present participle (phrase)

The verbs of this pattern are followed by a noun, pronoun and a present participle, e.g. Firms have found themselves advising (FRT subcorpus), exchanges have found themselves having (FRT subcorpus).
[VP19C] \( S + vt + \text{noun/pronoun/possessive + ing form of the verb} \)

These verbs are followed by a noun/pronoun plus a present participle or gerund, depending on whether we have a noun or a possessive preceding, e.g. *Failure may lead to the firm closing out* (FRT subcorpus).

[VP22] \( S + vt + \text{noun/pronoun/gerund (DO) + adjective} \)

These verbs take a direct object and an adjective, expressing result or manner, e.g. *The exchange considers the specific floor practice undesirable* (FRT subcorpus), *the court may hold a rule invalid* (FRT subcorpus), *you have set page breaks intact* (CMT subcorpus), *we shall keep the voltage constant* (EMT subcorpus), we should choose the *potential as a vector* (EMT subcorpus).

[VP23] \( S + vt + \text{noun/pronoun (DO) + noun (phrase)} \)

In this pattern the predicative adjunct is noun, e.g. *we may call a ring current a magnetic dipole* (EMT subcorpus), *we shall call it a surface charge* (EMT subcorpus), *the relevant transactions make someone a customer* (FRT subcorpus), *Reference to other business does not make someone a customer* (FRT subcorpus).

[VP25] \( S + vt + \text{noun/pronoun (DO) + (to be) + adjective/noun} \)

Verbs falling under this category are followed by a noun/pronoun and a complement after *to be*, e.g. *the firm finds itself liable* (FRT subcorpus).

Notes

1. The notion of subsystem was originally introduced by Harris (1968, p. 152) who drew a parallel between subsystems in Mathematics and subgroups in Language. He stated that: *"certain proper subsets of the sentences of a language may be closed under some or all of the operations defined in the language and thus constitute a sublanguage of it"*.

2. We do not include structure \( S + vt + \text{noun} \) in our discussion as it is too general to be related to any special language.

3. It should be noted that the percentages in Tables V and VI are estimated on the basis of how often we have a term in a certain syntactic position when this position is filled in by a lexical item and not by the total number of occurrences of a certain grammatical structure.

References


