Verbs of Science and the Learner's Dictionary

Geoffrey Williams
Université de Bretagne-Sud

This paper looks at how the verbs of science are displayed in the OALD and then compares them to a specialised corpus. The individual entries will be studied to see whether the scientific aspect is signalled, how the definitions are structured, and if implicit information is carried in the examples. The examples are analysed using Halliday's Systemic and Functional Grammar (SFG). The results in each section are compared with usage cases found in a subcorpus from the British National Corpus and in a specialised corpus.

1. Introduction

Since Barber (1962), it has generally been accepted that nominalization is a key feature of science writing as it permits the writer to stand aloof from the text, thus giving an impression of greater objectivity and authority. This rhetorical stance has lead to the semantic weight being placed on noun-forms and explains why traditional terminology has concentrated almost exclusively on this aspect of scientific usage. The pre-eminence of the noun is also found in specialised vocabulary lists, such as that of the *Oxford Advanced Learner's Dictionary* (OALD).

In addition to a 3,000-word keyword list (R99), the OALD also gives a list of 250 scientifically orientated vocabulary items (R115). Whilst this list may be far from perfect (Williams 2006), it does give an interesting breakdown as to what a major dictionary sees as significant general scientific usage. It is clear that nouns dominate, followed by adjectives and the verbs. Given the lowly place of the verb it will be interesting to see what verbs survive and how their role in scientific usage differs from that of general language. The premise is that although the OALD, like all learner's dictionaries, aims essentially at general usage, they could equally be well be used by students studying English for Specific Purposes who would be seeking assistance with encoding as much as decoding.

In this paper we shall look at how the verbs of science are displayed in the OALD and then compare with a specialised corpus. The individual entries will be studied to see whether the scientific aspect is signalled, how the definitions are structured and what implicit information is carried in the examples. The examples are analysed using Halliday's Systemic and Functional Grammar (SFG), supplemented with role analysis adapted from FrameNet. The results in each section are compared with usage found in a subcorpus from the British National Corpus and in a specialised corpus. The aim is not to criticise the OALD or any existing dictionary, but to see how individual entries may be expanded or adapted to suit learners coming the sciences,

2. The Oxford 250 and its verbs

The Oxford 250 is in fact a list of 266 word forms, once derived forms and spelling variations have been taken into account. Forty four verbs are listed of which 27 are purely verbal in nature whilst the remaining 17 are classed as either verb and noun or noun and verb. Thus *attribute* is classed as V, N and *bond* as N, V (table 1). This is obviously based on frequency in their general sense rather than their importance to scientific discourse. In this paper we shall concentrate on these V, N and N, V forms.

attribute	v, n	bond	n, v
coordinate	v, n	clone	n, v
discharge	v, n.	cluster	n, v
probe	v, n.	extract	n, v
scan	v, n.	filter	n, v
stain	v, n.	orbit	n, v
yield	v, n.	pump	n, v
		sequence	n, v
		shield	n, v
		stem	n, v

Table 1. The V, N et N, V verbs in the OALD

If we start with the V, N words, that is those whose verbal form is more frequent in general language, we find that only three are specifically signalled as being scientific and even then it is the noun form of *probe* that is signalled as being of scientific importance (table 2). In five cases the technical aspect can be found implicitly in either the definition or the example. This is true in all seven cases for the noun forms, which would imply that the real scientific nature of these words lies in the noun.

Headword	Signalled	Verb usage	Noun usage
Attribute			Quality, feature
Coordinate			Graph
Discharge	technical	Release force, power	Implicit – action of ~
Probe	technical	Implicit - instrument	Device, tool
Scan	computing	Program, memory	Program, memory
Stain		Microscope	Implicit - medical
Yield		Research, information	Crop yield

Table 2. The V, N forms in the OALD with their signalling label or words in the definition or example that imply a scientific usage.

A similar picture can be found when looking at the N, V forms (table 3) although this time 7 of the 10 receive a label. However, the presence of *cluster* seems odd in that it is only labelled under phonetics, which is not generally considered as a science as defined in the OALD context. The use of this word in biology and computing is not signalled. In all other cases there is a clear indication that these words function as both nouns and verbs within the sciences.

Headword	Signalled	Verb usage	Noun usage
Bond	Chemistry	Chemical compound	Implicit - atoms, molecule
Clone	Biology	Cells	Implicit – cells
Cluster	Phonetics		Consonants
Extract	Technical	Dentistry - force, effort	Implicit - substance
Filter	Computing	Implicit –program	Program, device
Orbit		Curved Path	Curved Path
Pump		Machine	Make flow
Sequence	Technical, Biology	Identify order	Implicit - Set, order

Shield		Protect, rays	Protect, ozone layer
Stem		Stop flow	Plant

Table 3. The N, V word forms in the OALD

It is clear that the definitions and examples play a key role in demonstrating the scientific nature of the word. In the OALD verbs are generally defined using the infinitive with the following model:

[headword inf] [is] [to do something]

for example:

[to stain] [is] [to leave a mark that is difficult to remove on sth]

Here, the scientific aspect is found in the example with the word *microscope*. In the case of *scan* the keywords are located in the definition itself:

[to scan] [is] [to examine a computer program or document in order to look for a virus]

This is then reinforced with the example: "*This software is designed to scan all new files for viruses."*

The examples either reinforce or supply the scientific aspect, this being particularly important in a learners dictionary. In some cases, for instance *bond*, *clone*, *cluster*, *filter*, the editor has decided that the noun usage is sufficiently clear to not warrant an example. However, for *bond, clone* and *filter* a link is operated by reception of definitional elements between the signalled noun and the associated verb form, which does have an example. Thus, if we take *bond*, the definition declares one sense of the noun to be:

CHEMISTRY
8[C] the way in which *atoms* are *held together* in a chemical compound

The first verb form picks up part of the definition with the key word atom being in an example.

1bond (A and B) (together) | bond (A) to B to *join* two things *firmly together*; to join firmly to sth else:[vn]
[v] The *atoms* bond together to form a molecule.

Thus we have decoding assistance for the noun, but also encoding assistance for the verb form. In the other cases illustrative examples can also be found, although their scientific or technical nature is not always easy to see, as can be seen with the "technical" and "formal" examples for *extract* for which the first example is clear enough, but not the second:

The dentist may decide that the wisdom teeth need to be *extracted*.
He rifled through his briefcase and *extracted* a file.

In the case of *stem*, the verb form is not labelled as technical, although the example relies on the context word *wound* to supply this aspect as in the phrase "The cut was bandaged to *stem* the bleeding". It is of course possible that the technical usage is purely linked to the noun form in the context of plants with the link to stemming blood not being considered as technical in this instance.

In terms of tense, various forms are found; active, passive, imperative, infinitive and gerund. This has its importance as different verbs have their preferences within a textual environment; and also affect the thematic construction of a sentence. Given that the sciences are supposed to privilege the passive, it could be seen as surprising that there is only one example.

In order to look more closely at the grammatical and textual implications of the definitions we shall apply the analytical tools of Systemic and Functional Grammar (Halliday 1994) taking as examples only those verbs that have an example on the basis that a student studying sciences would be seeking encoding, as well as decoding, information.

3. Systemic and functional analyses

Systemic and functional grammar (SFG) has been chosen for two reasons; first it is a textual grammar that allows lexicogrammatical choices to be placed in a wider environment than the sentence and secondly it is the most widely used grammar in English for Specific Purposes, precisely the community for which we propose to adapt general dictionaries.

SFG distinguishes three broad metafunctions in the analysis of language; these are the textual (clause as message), the interpersonal (clause as exchange) and ideational (clause as representation) functions. The first function includes the all important aspect of theme-rheme, vital for the structuring of information, the interpersonal deals with communicative acts and aspects such as modality, whilst the last, the ideational function, concerns content that may be expressed in terms of roles. In looking at the last element we shall also draw on the terminology of FrameNet in that it is far richer than that proposed in SFG analyses.

The first examples (table 4) are all active forms and therefore easy to analyze at the thematic level as the subject corresponds to the theme. However, the ideational aspect as outlined in SFG (Halliday 1994) is more problematic.

Verb		*NP*	*V*	*NP*	*PP*
Grammatical		*Subject*	*Predicator*	*Complement*	*Adjunct*
Textual		*Theme*		*Rheme*	
Ideational	*Role*		*Pro: Mat*	*Goal*	
Bond	*Actor? Item*	The atoms	bond		together to form a molecule
Probe	*Actor*	The doctor	probed	the wound	for signs of infection
Probe	*Instrument*	The searchlights	probed	the sky	
Pump	*Instrument*	The heart	pumps	blood	around the body
Shield	*Force/Instrument*	The ozone layer	shields	the earth	from the sun's ultraviolet rays
Yield	*Instrument*	The research	has yielded	useful information	

Table 4. Ideational analysis of active forms

Following an SFG analysis, these verbs are classed as being material processes, which would normally require an actor doing something to something else (goal). The actor can also be replaced by an instrument or force. These verbs may be transitive or intransitive.

Our first example is transitive, and problematic in that there is no actor as such, the binding being caused by an unnamed force. In the BNC, the most common subject of *bind* is indeed atom. In FrameNet, *bond* is part of the attaching frame, which usually has a human actor, or agent with what is bound being the item. In scientific usage, however, the item is the subject as no agent is present or possible.

Probe should be easy as we have a human actor in the first case who wields an instrument that has the characteristics of the noun *probe*. In the second example, the subject ius the instrument. It could be suspected that in scientific usage, where actors are rare, there is a transfer of meaning from noun to verb which simply places the searchlights as an instrument that acts as a type of probe. This could be seen as a case of collocational resonance (Williams 2007 and 2008 forthcoming) where part of the meaning of a word is carried over to a new textual environment. In this case 'a probe' is a long thin instrument with the instrument role of the verb retaining the

long thin characteristic as with a long thin shaft of light. Framenet, on the other hand, sees *probe* as primarily a mental process with *cognizer* as subject within a *scrutiny* frame. This again illustrates the potential variance with scientific usage that would need clear signalling to the learner, but, as we shall see later (section 4), it is the FrameNet analysis which may be closest to real scientific usage.

For both *pump* and *shield* what we have is an instrument or force carrying out the process. *Pump* and *shield* follow a similar to *probe* in that the accepted process is that the actor *pumps* a liquid using an *instrument* whilst for *shield* a force or instrument provides a *shield* or *protection*. As with probe, the actor is rarely present so the subject becomes the instrument carrying over certain characteristics of the noun.

For *yield*, none of the analyses in the current version of FrameNet really suit as the research is the "revealer" not the "creator" of the information. In many ways the closest synonym would be *reveal (evidence)* wherein the *instrument* is classed as a *support* with the *results* or *outcomes* of the research, when interpreted by the *researchers,* providing information. The activity of research is thus an instrument for finding information. This is confirmed by the set of subjects that *yield* will take in the BNC as these all require the intervention of a human researcher: *analysis*, *approach*, *method*, *project, strategy, study.* To understand the use of the verb *yield* in this context the learner would need access to this set of nouns.

Textual		*Theme*	*Rheme*		
			Finite	*non-finite*	
Ideational	*Role*		*Mood*	*Pro: Mat*	*Goal*
Clone	*Actor*	A team from the UK	were	to clone	an animal
Scan	*Instrument*	This software	is designed	to scan	all new files for viruses
Stem	*Goal*	The cut	was bandaged	to stem	the bleeding
Pump	*Instrument*	The engine	is used for	pumping	water out of the mine
Discharge	*Force*	Lightning	is caused by	discharging	electricity

Table 5. Material processing and non-finite clauses

Such hypotactic structures obviously demand more from the learner who has to recognise the structure if it is to be reproduced. Four of these examples are quite clear as there is an actor, instrument or force at work. Here, it is the verb *stem* that poses a difficulty as the theme refers to the wound itself, the goal of the verb *probe*.

To sum up, if entries are to be useful to someone producing language, especially if it is in a scientific discourse, then it will be necessary to reveal the normal phrasal patterns and frames of these verbs within a scientific context. It will also be necessary to see whether these verbs do actually occur regularly in scientific discourse, and if they do not what verbs do. To do this we shall look first at the British National Corpus using Sketch Engine, then a specialised corpus.

4. Verbs in the BNC and specialised corpora

The problems with the BNC are known. It is no longer a recent source and its subsections cannot really be considered as representative of a register or genre (Aston 2001). However, it remains the only widely available balanced corpus for English and, as available through Sketch Engine and the XML format of the BNC World Edition, is still a useful reference for general lexicographical purposes. The other source used is a corpus in the making, SCIENTEXT (*http://w3.u-grenoble3.fr/lidilem/labo/scientext/*) that aims to be a comparative reference corpus of academic usage in French and English. In this study we shall refer to the English corpus that is currently under construction in Lorient. This means that we are potentially looking at two

levels of text; very general scientific usage as shown in the BNC and the full research paper used for SCIENTEXT. Whilst the general language user is likely to only encounter popular science, the ESP student will quickly come into contact with the research article. The dictionary writer cannot be expected to cope with both, but the teacher could show how to move between genre. This is certainly better than the blind faith in the bilingual that is often encountered with both ESP students and scientists in general.

We shall start with the four verbs dealt with at the end of the previous section, namely *probe*, *pump*, *shield* and *yield*.

Probe (table 6) occurs 240 times in the science subdivision of the BNC as compared to 727 for the corpus as a whole. These figures are not exact as this word obviously posed problems for the tagger with many entries under verb being noun forms. In order to partially get round the tag problem the overall figures have been extracted using the World Edition of the BNC and the sense investigation carried out with Sketch Engine.

A glance at the concordance confirms the heterogeneous nature of this subsection with a variety of 'lightly' scientific contexts that include anything to do with nature, including gardening. Forty of the forms are non-finite clauses using the infinitive. These seem to indicate two main senses; exploring with an instrument and investigating, the latter being the mental process described in FrameNet.

The first sense seems to correspond to that shown in the OALD with an instrument being wielded by an actor. These are inevitably very lightly scientific, if at all. The most used scientific sense seems to correspond more closely to the FrameNet scrutiny frame, hence a slightly different role analysis. As ever in research papers, the investigator, *cognizer* in FrameNet terminology, is absent.

Textual		*Theme*	*Rheme*		
			Finite	*non-finite*	
Ideational	*Role*		*Mood*	*Pro: Mat*	*Goal*
Clone	*Actor*	They also	use sticks or grass stems	to probe	for food
FrameNet	*Role*	*Means*		*Scrutiny frame*	*ground*
Clone	*Instrument?*	Clearly more experiments	are required	to probe	the role , if any , of the C-terminal helix
Scan	*Instrument*	Electrochemistry	is also being used	to probe	the electronic structure of the fullerenes

Table 6. Role analysis for BNC examples

A similar situation is found with probing as non-finite in the applied sciences subcorpus of the BNC with sentences as:

> Ions are fine for *probing* electrically conducting materials
> researchers that use d3CA for *probing* DNA-protein and DNA-drug interactions
> was isolated from the EMBL4 gene bank by *probing* with the recombinant plasmid pGW1

In all three cases the verb could be replaced with a synonym such as *investigate*. We are, however, in a very borderline situation as the means is being used as a probe in the scientific sense of an entity used to as an investigatory instrument. We are far from the thin instrument in these contexts although the exploratory aspect remains, The "thin instrument" sense does occur in the natural sciences as in:

thin beaks delicately *probing* the water for weeds

In this case the beaks belong to the birds and are hence an instrument wielded by an animate entity.

Finite usage of the verb is relatively rare in the scientific setting with the passive being the rule:,

Blots were *probed* with monoclonal antibody supernatant

Present tense usage is limited to semi-medical situations with the cleaning of wounds. In other words, the technical sense given in the OALD is primarily medical rather than scientific and does not reflect the reality of scientific usage, which is carried by the first sense illustrated in this dictionary, that of investigate.

The scientific usage found in the BNC is confirmed in SCIENTEXT (table 7)in which only 27 of the 510 uses of the item are verbal with 17 being passive and 4 infinitive. The sense is exclusively that of investigate.

Goal	Finite		Instrument
Protein extracts	were probed		with antibodies
Instrument	Finite	Non finite	Goal
The contrast variation method	enables us	to probe	the structure factor of each polymer

Table 7. *Probe* in SCIENTEXT

Pump (table 8)is less polysemic being firmly linked to the instrument sense and the material process. It occurs 973 times in the full BNC as verb and 191 times in the science subsection. The passive (93/191) and non-finite usage (76/191) dominate. In SCIENTEXT, only one non-finite example is found. What we have in the BNC is a very general usage wherein a force, in 16 cases the heart, pumps something. The most scientifically oriented occurrences in the subsection tend to show the action of pumping through the infinitive. It is clear that the ESP user needs to be introduced to such clause structures, as well as to the notion that a pump is not just to move a liquid, but to apply pressure,

Force	Non-finite	Goal
which uses steam ejectors	to pump	highly active liquor around the plant
a transverse AC electric field	to pump	energy into ions in motion in that field

Table 8. *Pump* in SCIENTEXT

Shield (table 9)also appears primarily in the passive in both the BNC and the SCIENTEXT corpus. Here again we adopt the Framenet terminology for greater clarity to show that the set of "assets", that which is to be protected, are shielded by a form of protection, as in the first example or from a set of dangers. The preposition that helps the learner to differentiate the protection from the danger.

	Asset		Protection/Danger
BNC	peripheral computer equipment	are shielded	or enclosed in magnetic toroids
SCIENTEXT	the cups and growth media	were shielded	from the spray solution

Table 9. *Shield* in the BNC and SCIENTEXT

The last of the verbs is *yield* (table 10) with an occurrence of 2,157 in the BNC as a whole for only 37 in the science subsection. In the specialised corpus it occurs only 12 times out of 200 occurrences as a verb. *Yield* occurs primarily in the active, often with a modal verb. *Yield* is frequently linked to crops, but the keyword *research* is also found in the SCIENTEXT corpus. It is apparent that two senses must be distinguished; that of crops and that of the limited set of research and methodology nouns introduced in section 4.

	Instrument		theme
BNC	The barley	yielded	more than 4t/acre last year
SCIENTEXT	further research […] continues	to yield	such exciting results

Table 10. *Yield* in the BNC and SCIENTEXT

What is immediately apparent from this research is that these verbs do not have the same patterns as those shown in the OALD. It could be useful for the learner's dictionary to have a wider range of examples, but in practical terms the teacher needs to use the dictionary as a starting point and then move onto structures used in more specialised texts, maybe through the use of a corpus. Another important conclusion is that the sense signalled as being technical or scientific may not be that which dominates in the sciences, as in the case of *probe*.

A third finding is that the verbs investigated are far from frequent in scientific usage compared to their noun-forms.

It is true that we have dealt only with a small subset of verbs which have a nominal homograph. However, comparison with the 27 remaining verbs of the OALD list with the SCIENTEXT corpus shows them not to be particularly frequent. It is obvious that the current study needs to be enlarged to investigate their scientific usage. However, their lack of frequency does lead to the question as to what verbs are used in the sciences. Part of the answer to this can been found in looking at lists obtained from two very different subsections of the SCIENTEXT corpus.

5. The verbs of science

SCIENTEXT is a collaborative project with research teams from three French Universities, Chambéry, Lorient and Grenoble. The project consists in the building and analysis of a large comparable English - French corpus of scientific and academic language. The resulting corpus will ultimately be accessible on-line.

The English corpus is being built at the *Université de Bretagne Sud, Lorient*, and aims to be a general corpus of scientific English drawn from a variety of specialities. Work on the collocationals patterns in a corpus of parasitic plant biology is part of an ongoing study published elsewhere, (Williams 2002|2007). Here (table 11) we compare two interdisciplinary subcorpora; *polymer sciences* and *parasitic plant biology*. Both are interdisciplinary in that a number of domain specialities are involved, physics and material sciences in the first, plant biology and biochemistry in the second. This will allow us to isolate verbs that are common to both and which, whilst being classable as general language have precise usage in the sciences. The first column gives the common verbs found in both sub-corpora amongst the 30 most frequent verbs. The second and third show the most frequent verbs of each sub-corpus. All three columns are by order of frequency.

Common	Parasitic Plants	Polymer
use	use	use
show	show	obtain
obtain	find	show
observe	grow	form
find	observe	observe
study	develop	increase
suggest	suggest	present
contain	study	prepare

determine	obtain	measure
make	identify	contain
see	describe	give
allow	indicate	report
base	occur	study

Table 11. Most frequent verbs in two subcorpora of SCIENTEXT

As can be seen there is little that is domain specific. These are relatively general words, mostly concerned with observation. but their usage can be specific to a domain, even for such general words as *use* as has been shown by Richalot (2003). Within both sub-corpora used here the passive dominates (table 12), but what the ESP learners will need to see are the various arguments and roles that are associated with the verbs, a task where the rich analyses of FrameNet can be associated with a SFG clause structure analysis of examples. The learner's dictionary remains a starting point, the corpus analysis will provide the specificity.

Agent		Instrument	Purpose
Sorghum and pearl millet	were used	as host plants	
Genomic DNA [...]	were used	for the Polymerase Chain Reaction (PCR) technique	to determine the genetic polymorphism
Lacmoid blue and aniline blue methods	were used		for localization of callose on the sieve plates

Table 12. Using *use* in SCIENTEXT

6. Conclusion

The analysis of this small number of verbs shows clearly the problem that a general language dictionary faces when tackling languages for specific purposes. However good a specialised vocabulary list may be, it is inevitably based on a very broad definition of science. It can also suffer from reliance on a non-representative subcorpus and an insufficient analysis of scientific usage. In a teaching context this can be overcome quite easily by adapting the examples to the realities of usage. However, it can be hoped that access to a balanced corpus of scientific English will also allow the creation of word lists that reflect genuine scientific usage at a general level. Through SCIENTEXT, we hope to produce a general lexicon of scientific verbs and show their specificities by combining SFG and FrameNet analyses. In this way it might be possible to improve current dictionary coverage of sciences and also to provide a methodology whereby the learner user can benefit from the clarity and depth provided by learner's dictionaries whilst adapting the content to areas of more specialised usage.

References

Aston, G. (2001). "Text categories and corpus users : a response to David Lee" [online]. *Language learning and technology* 5 (3). 73-76. *http://llt.msu.edu/vol5num3/pdf/aston.pdf*. [Access date: 14 November 2007].

Barber, C. L. (1962). "Some Measurable Characteristics of Modern Scientific Prose". In Swales, J. (1985). *Episodes in ESP*. Hemel Hempstead: Pergamon Press. 3-14.

Halliday, M. A. K. (1994). *Introduction to Functional Grammar*. 2nd ed. London: Edward Arnold.

Kilgarriff, A.; Tugwell, D. (2002). "Sketching Words". In Corréard, M. H. (2002). *Lexicography and Natural Language Processing: A Festschrift in honour of B. T. S. Atkins*. Stuttgart : EURALEX. 125-137.

[OALD]. *Oxford Advanced Learner's Dictionary*. 7th ed. Oxford: Oxford University Press, 2005.

Richalot, J. (2003) *Approche énonciative du contexte d'occurrence des candidats termes: constitution et exploitation d'un corpus d'angalis en langue de spécialité*. Unpublished PhD thesis. Université de Lyon III.

Williams, G. (2006). "Advanced ESP and the Learner□fs Dictionary: Tools for the non-language specialist". In Marello, C. (ed.). *Proceedings of the 12th EURALEX International Congress*. Turin: Université de Turin.

Williams, G. (2007). "De l'architecture des sources à l'architecture de l'entrée : le rôle du corpus". In Dotoli, G. (2007). *L'Architecture du Dictionnaire Bilingue et le Métier du Lexicographe, Actes du Colloque International de Capitolo-Monopoli, 16-17 avril*. Fasano: Schena. 39-53.

Williams, G. (à paraître 2008). "The Good Lord and his works: A corpus-based study of collocational resonance".□In Granger, S.; Meunier, F. (2008). *Phraseology : an interdisciplinary perspective.* Amsterdam: Benjamins.

Williams, G. (2007). "In search of representativity in specialised corpora: categorisation through collocation". In Teubert, W.; Krishnamurthy, R. (eds.). *Critical Concepts in Corpus Linguistics*. London: Routledge. Vol. 3. 324-346.