Ahmad, K. (2000). 'Neologisms, Nonces and Word Formation'. In (Eds.) U. Heid, S. Evert, E. Lehmann & C. Rohrer. The 9th EURALEX Int. Congress. (8-12 August 2000, Munich.). Vol II. Munich: Universitat Stuttgart. pp 711-730. (ISBN 3-00-006574-1).

Neologisms, Nonces and Word Formation

Khurshid Ahmad
Department of Computing, University of Surrey
Guildford, Surrey, United Kingdom, GU2 7XH.

Abstract
Neologism dictionaries celebrate the influence of science and technology on language – how scientists and technologists add to the stock of words, and occasionally, add new grammatical structures. There are two claims made in the literature on neologisms: First, that borrowings, from classical languages and 'prestige' languages are widespread in the science and technology literature. Second that affixation is an ephemeral mechanism for forming new words. Bold claims? Let us look at some of the evidence for and against these claims.

Preamble

Lexicographers have a curious relationship with neologisms and neologists: Sometimes they will go out of their way to embrace a new term; for instance, the OED acknowledged the work of the physics nobel laureate, Murray Gell-Mann, and attributed to him the coinage of the term *quark* - the most recent building block of matter which started life as a superordinate term for *up*, *down* and *bottom* quarks. And, Gell-Mann in turn was inspired by a James Joycean character *Muster Mark*, an elusive character who had at least three personalities - ' three quarks for Muster Mark' (Gell-Mann 1997). At other times lexicographers assume the defence of the language whose lexicon they are compiling and become quite possessive about their language. This defence, and possessiveness, shows itself by their unease with words/terms being imported from another language. The role of language planners, comprising national language (planning) bodies like the Academie Française and Real Academie, or the various terminology standardising bodies (like Union Latine) which work under the umbrella of standardisation organisations, is an ambiguous one insofar as neologisms are concerned. On the one hand, the national language bodies, aided and abetted by the national press (McMahon 1994:174), are quite jealous of changes in meaning and in lexical inventory: in Britain we frequently complain about "Americanisms" from across the water. On the other hand, the march of science and technology, the emergence of multinational sporting events and the globalisation of food and drinks, forces the hand of language planners to let in foreign words and also syntactic structures of other languages. In France, the Academie Française tends to regulate the influx of loan words which results in quite curious situations when *broadcasting* has to be called "télédiffusion", *hovercraft* as "aéroglisseur" and *planning* as "planigramme" (Picone 1996:282). In both cases, the planning bodies have to deal with neologisms in one form or the other.

Neologisms are an interesting phenomenon in that their emergence demonstrates the capability of language to undergo and sustain change, and its capability of deflecting negative intrusion from other languages and cultures. Many authors, including Crystal, describe neologisms as "nonce" words in that of the many neologisms created, adapted, mutilated, very few survive. A nonce word is 'a linguistic form which a speaker consciously invents or accidentally does on a single occasion [……] Nonce formations have occasionally come to be adopted by the community – in which case they cease, by definition, to be 'nonce' (forms used 'for the (n)once') and become neologisms' (Crystal 1997). Crystal reinforces an earlier statement from Quirk et al (1985) that 'the vast majority of such new formations remain uninstitutionalised attempts at lexicalisation'.

The Oxford English Dictionary suggests that *neologist*, an early 19th century word, had two senses: First, a neologist was a person who invented or used new words or forms (c. 1785). Second, a neologist was a person who was prone to rationalisation in theology or religious matters (c. 1827). The first sense of the word neologist, and by implication *neologism*, is still with us. The second sense of the term neologism is now obsolete except for the fact that scientists and technologists are the new rationalists here. They wield as much influence now, if not more, than their religious counterparts in the 19th

century. It is the scientists as technologists who attempt to rationalise our experience of the world around us in written language by using new words or forms or by relexicalising the existing stock. This can be attested not only by the acknowledged influence of science and technology on language, but also by the growth in the publication of specialised dictionaries and in the dissemination of terminology through the World-Wide Web. The growth of terms in each successive edition of specialist dictionaries attests to the role of scientists, both experimental and theoretical: *Electronics*, *Computing* and *Physics*, and *Linguistics*. (See Table 1).

Dictionary (Publisher)	***Year of Publication (Edition); No. of terms***	***Year of Publication (Edition); No. of terms***
Electronics (Penguin)	1998 (3rd Ed.); c.4700	1988 (2nd Ed.); c.3800
Computing (Oxford)	1996 (4th Ed.); c.6000	1990 (3rd Ed.); c.4300
Linguistics and Phonetics (Blackwell)	1996 (4th Ed.); c.6000	1991 (3rd Ed.); c.5300
Physics (Oxford)	1996 (3rd Ed.); c.4000	1990 (2nd Ed.); c.3000

Table 1: Lexical growth in specialist subjects

In this paper, we will look at word formation – as one particular source of neology -in the specialist languages of science and technology (and consider linguistics as a science of language). We will look at the 'institutionalised attempts' at (re)lexicalisation, by learned societies and publishers of learned texts, especially by borrowing and affixation. We examine borrowing from classical languages in newly emergent subjects like computing, and affixation as a process for coining new terms.

Neologisms and lexical change

The development of the *New Oxford English Dictionary* involved initially setting up a target of adding 'three thousand of the most significant words and meanings to the integrated edition [*OED + Supplements*]', and subsequently 'the monitoring of new words and the making good of deficiencies in the dictionary's coverage of the English vocabulary' (Weiner 1987:39). Now many dictionary publishers, at least in English, claim to *monitor* (a relexicalised word itself) changes in their language.

Global movements like the 1970's environmental awareness campaigns have contributed to the sanitization of goods, products and services which in their older forms may show a lack of sensitivity to environmental concerns but when prefixed with 'eco-' become okay: warriors became *eco-warriors*, and the disruptive tourists could go on *eco-tours* and help *green* the Third World. We now have *ethical foreign policies*, *sustainable development*, *stakeholders* and *Scientists Against Nuclear Weapons* (*SANE*). Compilers of neology collections in a number of cases now seek to link the process of nonce formation and neologisms to catastrophic events and to global movements. Thus for Ayto (1999) the financial crisis that hit Western Europe and the USA in 1987 put an end to the 'exuberance of inventiveness' of the stockbrokers, money men and junk-bond dealers.

Some lexicographers suggest that there are three broad 'source types' of neologisms in dictionaries as well as in corpora: first, neologisms formed by the addition or combination of elements, especially compounding, affixation, blending and acronymization; second, neologisms formed by reduction of elements, namely, abbreviations, backformation and shortenings; and, third, neologisms that are neutral with respect to addition or reduction: semantic change, coinages, conversion or loans. For John Ayto, neologisms formed by addition 'survive relatively less well than types formed by reduction […] or than types that are neutral with respect to addition or reduction […] [b]ack formation is an exception to this tendency' (Ayto 1995:187). It should be noted that prefixes and suffixes account for about two-fifths of new words in *Merriam Webster* and around three-fifths involve compounding; the rest include borrowing, conversion and backformation.

McMahon, in her largely diachronic study of semantic change, talks about *lexical creativity*: 'the formation of new words using a language's own resources, including productive morphological processes and compounding' (1994:174). For her, one of the identifiable aspects of language which allow semantic change to occur is the fact that "words are typically polysemic" and hence "can lose or gain meaning relatively easily [...] and do not have to lose an earlier sense to gain a new one". The word *atom* is a good example - prior to 1910 it was an indivisible unit of matter and since then, through

relexicalisation, it has gained constituents which, in their own right, can be subdivided down to the new indivisibles, the *quarks*. However, the word *atom* is still used in other specialisms like computing and linguistics, and also in general language to refer to objects which do not have structure or constituents. If and when computing folk catch up with the high-flying physicists, we will have *quarks* (instead of atoms) to refer to an arbitrary string of characters.

A note on word formation

Grammarians, with their emphasis on making generic and abstract statements about the behaviour of phrases, clauses and sentences, and lexicologists, with their focus on specific statements about individual lexical items, appear to be at the polar ends of a continuum. However, in the area of *word formation*, both grammar and lexicology 'share a common ground: where generalisations, as in grammar, are appropriate, but where idiosyncrasies of individual units [lexical items] are also described' (Quirk et al. 1985:1517). For Quirk et al (1985), a typical language user – the 'person-in-the-street' – is usually 'passive' in word formation unless he or she is a 'poet' or an 'experimental scientist'. The early English lexicographers, from Samuel Johnson on to the 19th century dictionary compilers, relied largely on religious writings and leading playwrights/poets for the supply of new words, or for the relexicalisation of extant words. Jespersen notes that in turning the pages of recent dictionaries 'one is distracted by the frequency with which Shakespeare's name is affixed to the earliest quotation for words or meanings' (1938/1962:211). The typical 'person-in-the-street' was, and is still, regarded as *passive* in word formation (Quirk et al 1985, Aitchison 1991). From the late 19th century, scientists and technologists have asserted their authority, and earned in roughly equal amounts the respect of parts of society by their inventions and opprobrium for science-made disasters. The influence of scientists and technologists can be measured by the fact that the many of the neologisms they coin became acceptable to lexicographers, who are usually rather conservative in their outlook. The 20th century saw the rise of marketing men and the advertising industry: their authority asserts itself indirectly through 'campaigns' on behalf of their clients, and some of the advertising/marketing coinages have made their way into major dictionaries. Indeed, some semanticists are of the view that the study of the *creative* neologisms of these creative people will contribute to our understanding of language itself (see, for instance, Lehrer 1996). We will, however, focus on science and technology.

Borrowing and neo-classical formation (Quirk et al 1985), and, in some cases, "pseudo-Classical neology" (Picone 1996) have always served scientists and technologists well. The use of Greek or Latin words had dominated science and technology literature written by the Arabs, who admixed some Indo-Aryan languages as well. In the 19th and 20th centuries, scientific literature written in English, German and French, and other Indo-European languages, shows extensive neo-classical formations, indicated by number properties (mono, multi, bi, etc.) and by the assimilation of a number of words from Latin, Greek and Arabic. However, the adapted words show no trace of their origins: the Arabic *al-kimiya* (miracle or magic) became chemistry, and *al jabar wal muqabla* became *algebra*, and the rest (of the assimilated words like *chemical, chemistries, algebra, algebraic*) are, as they say, history. However, there has also been borrowing amongst the so-called modern languages and Picone tells us that "it is the French who forged *biologie* (1802), *sociologie* (1830), *automobile* (1860), *cinématographe* (1895) and *radioactivitie* (1896), words whose English equivalents betray no trace whatever of their [French] pedigree" (Picone 1996:291). This assimilation, by the way, works both ways in that French specialist literature made a number of "integral borrowings" which reflect the initial adaptation of prefixes like "self" as in *self-defence* (1869), *self-induction* (1881) and *self-portrait* (1925); and within a few years of this pseudo-Classical neology, these terms were assimilated as *auto-induction* (1890), *auto-defence* (1896) and *auto-portrait* (1928). Picone calls this *juxtapositional neology*: Elements of an expression, sometimes by virtue of repeated use, were simply frozen in their naturally juxtaposed position (1996:32). This "lexicogenesis" has manifested itself in the adaptation of terms like *surface-to-air missile* to the French *missile sol-air*, but, according to Picone (1996:263) we don't see "missile du sol et du l'air" or "missile entre le sol et l'air" in the French literature and subsequently the complex English term *anti-missile missile* was rendered simply *missile anti-missile* (Picone 1996:309). This borrowing from a perceived prestige language brings with it not only vocabulary, translated or original, but also sometimes contributes to the adaptation of syntactic patterns and morphological structures as well.

In the specialist literature of science and technology and, to a lesser extent in other enterprises like leisure and entertainment, there is extensive use of affixes and compounding, juxtaposed or not. It is the use of suffixes that results in the change of the word-class which is used extensively. For example, the lemma *react* can be suffixed to form *reaction* and then further modified to form *reactions*, and even more interestingly, *reactant* and *reactants*. Once the term *reaction* was created, then whole new branches of science and technology were formed around it beginning with *chemical reaction* and then going on to *nuclear reaction*. And, now in politics we have the *reaction of the masses* and *governments*. In these cases, the range of meaning has increased, in the first instance (*react* → *reaction*) and then there is a focus put on the meaning by the use of compounding (*chemical reaction* vs. *nuclear reaction*). A similar point can be made about *attract* → *attraction*; *gravitational attraction* vs. *electrical attraction* vs. *nuclear attraction*.

Quirk et al give us eight prefix categories, which might be broadly interpreted as semantic categories (1985:1540-46). Most of these categories are relevant for scientific and technical writing, and comprise prefixes to indicate "negatives" and include the commonly-used "non-", as in *non-metal* or *non-central*, and "in" as in *incomplete*. A negative prefix for showing the converse, is "dis-" as in *disorder*. The other categories include reservatives/privatives, for example the prefix "de-", as in *decentralised* and the prefix "dis-" as in *disinfect*. There are categories which include degree/size, like "sub", "super" and "hyper", and the category of orientation and attitude where the prefixes like "anti" (as in *antimissile*) are used. Finally there is the category of time which includes the prefix "pre", as in *prefix* or *prewar*, and "post" as in *postposition* and *postmodern*.

Ayto claims that, despite their overrepresentation in neologism dictionaries, new affixed words are the least likely to survive in the lexicon of a language: 'Affixation is easy, but maybe affixed forms are the leading disposables among modern neologisms: use them and throw them away […] such is the "nonce" feel of many of these forms […] that one might speculate about the extent to which they are true lexicalisations, and should validly be considered as part of the process of word formation' (Ayto 1995:186). We will look at this claim by Ayto by examining a case study on both prefixation and suffixation. However, before this, we would like to share with the reader some observations about how Classical and pseudo-Classical neology survives in recently emergent subjects, such as computing. This shows that Graeco-Roman influences on this branch of science and technology, and many others, is still alive and kicking.

Case Study I: Borrowing in (post-)modern times

Compilers of neologisms (Ayto 1999 and Green 1991 for example) have celebrated the influence of computers and communications technologies by including a number of terms from these subjects in their works. For Ayto, between 1940 and 1990, 'a small trickle of computer terminology [*electronic brain*, *hardware*] was to become a flood in the second half of the century' (1999:iv-v).

A closer look reveals that computer scientists have relexicalised words of Middle English origin (c.1150-1450), words like *circuit*, *digital* and *logic*. The lemma *compute*, in *comput*er or *comput*ing, rooted in the Latin *computare* (to reckon intensively), entered English in the 1630s. The modern variant *computer*, first used in the 1820s, referred to devices that computed by *weight*. Amongst the more frequently used terms in computer science currently are words that entered the language between 1550 to 1860 (all dates from the *SOED*, 1973): *algorithm* (c.1699), *automaton* (1625), *data* (c.1646), *hardware* (c.1555), *heuristic* (1860), *machine* (c.1599), *network* (c.1560), *procedure* (c.1611); *program* (c.1633).

Computer programs dominate our lives: from cash dispensers to flight control systems, the 'chip' in the food processor to electronically regulated flushing systems, computer programs are ubiquitous and pervasive; some are even called *ambient computing systems*. How has this term come into existence and spawned variants like *computer programs, computer programming, computer programmer(s)*? Table 2 shows the genesis of the compound term *computer program*. Over the relatively short period of a single century, the blend between the two different borrowings, used in very different contexts, is now seamless: one cannot in modern times think of a computer without a program, though Babbage's machines were just that, and a program cannot be thought independently of a computer. The post-

1950s computers, with their *stored programs*, and still reeking of the 19th century industrial revolution terminology of *core, mills* and *stores*, changed all that and made the two words inseparable.

COMPUTER n. 1897	PROGRAM (or PROGRAMME) 1895
Historical note: From the Latin *compute* and later from the French *computer* or computist, c.1631	Historical note: from the Greek *programma*, and later from the French *programme*; (pl. programmata c.1661) c.1633
1832 [computers] Medieval astronomical and calendrical calculations	1805 n. A descriptive notice issued beforehand of any series of formal proceedings – prospectus, syllabus, etc.
1842 [computist] One skilled in the computus or calendar; an accountant (obs.);	1820 n. (new obs.) A public notice
1848 n. An account	1831 n. A definite plan of any intended proceeding
1872 n. to compute by weight	1842 A public notice
1897 n. An [automatic] calculating machine.	
1946. Stored program(me) computer – a calculating machine capable not only of storing numbers to be computed but also the instructions. Computation could be carried out through analogue computers or digital computers.	
1996. A device or system that is capable of carrying out a sequence of operations in a distinctly and explicitly defined manner. A computer can have either a	stored program or wired program; the former can be altered much more easily than the latter.
executed by a computer in order to produce a desired behaviour.	1996. A set of statements that (after translation from programming-language form into executable form) can be
1990s. World Wide Web: a distributed information service that was developed at CERN. A distributed hypermedia system that is based on co-operating *servers* attached to a network, usually the Internet, and which allows access to *documents* (text, images, sounds) containing 'links' to other documents.	

Table 2: Lexicogenesis of *Computer Program(me)*

The neologisms coined by computer scientists appear to fall by the wayside rather quickly: nowadays nobody uses *psychons* - the basic units of thought as defined by McCulloch and Pitts (1943); terms like *memory organ*, *logical organ* and *arithmetic organ* coined by von Neumann (1958), whose 1945 design is still used in the bulk of the computing systems today, have made an appearance only in his book and (co-authored) papers. Marvin Minsky's *nomes* and *nemes* (agents in his idiosyncratic *The Society of Mind*) have largely only been used by Minsky himself to date. The neologism, *software*, appears to be a surviving exception.

Once modern computer scientists had consolidated their position, they began to appropriate terminology from other subjects for describing the human psyche and the brain. This process has continued to date and marked the emergence of post-modern computing. Two new terms characterise this appropriation*: Artificial Intelligence (AI)* and *Neural Computing*. Artificially intelligent programs and artificial neural networks are two strands of post-modern computing; AI borrows and reformulates terminology from epistemology, the neural networks community from neurobiology; both AI and neural networks attempt to reformulate terms in psychology. AI is an abbreviation for two Old French words - *artificiel* and *intelligence*; *neural network* is the combination of the Greek *neuron* (nerves or sinew) and the French *net*.

The existence of such inseparable compounds based on classical neologies is also pervasive in subjects like nuclear physics. Consider, for instance, the lexicogenesis of *atomic nucleus*: *atom* and *nucleus* were two distinguishable words up until the early 19th century. *Atom* and derivatives were used in philosophy (c. 1650) and *nucleus* in biology (c. 1820). The end of the 19th century shows the interest of physicists in the oxymoronish concept of *atom* (hitherto indivisible) with a constituent structure. The analogy of the solar system was used: an atom was like a solar system, with nucleus (read the *Sun*) and the electrons (read *planets*). Later in the first two decades of the 20th century we see the emergence of the compound *nuclear atom*, and, the more frequently used, *atomic nucleus*. Like the *computer program*, one cannot imagine the *atom* without the *nucleus*.

ATOMIC Atomicus – Latin	NUCLEUS (OED) Nucleus – nut, kernel, inner part
1692 Of or pertaining to atoms	1704 A more condensed portion of the head of a comet
1678 Concerned with atoms	1727 A supposed interior crust of the earth
1691 Adhering to the atomic philosophy	1869 A central or thing around which other parts or things are grouped, collected; 1829 (Botany, Zoology). That which forms the centre for some aggregate or mass
1809 Minute	1869 (Archaeology) 3c. A block of flint or other stone from which early implements have been made.
1901 Each atom might consist of …	one or more positive suns … and small negative planets (Perrin 1901)
1903 In an atom, [electrons move in one or more rings around]	a central body, much like Saturn and its moons (Nagaoka 1903)
1906 An atom comprises [electrons and]	'corpuscles', the number of the corpuscles is of the same order as the atomic weight of the substance.
1911 The atom contains	a central charge distributed through a very small volume … that the value of the central charge for different charge atoms is approximately proportional to their atomic weights (Rutherford 1911:687-688)
1913 […] every [atomic] system consist[ing] of electrons and positive nuclei	
	1932 Nuclei are made up of neutrons and protons (Heisenberg 1932)
[no ref. To atom!]	1949 [ ]Electrons, neutrinos and various types of mesons [ ] play an important role in the transformation and the attractive forces that occur between nucleons, thus securing the stability and even the existence of the composite atomic nuclei (Gamow & Critchfield 1949)
Becomes ATOMIC NUCLEUS in 1973 (OED)	
4. The positively charged central constituent, consisting in general of protons and neutrons, of the atom, comprising nearly all its mass but occupying only a very small part of its volume.	
NUCLEUS (of atom) 1996	
The central core of an atom that contains most of its mass. It is positively charged and consists of one or more nucleons (protons or neutrons) […]. The *simplest* nucleus in the hydrogen nucleus. […] The most massive nucleus […] is Uranium-238. (Isaacs 1996).	

Table 3: Lexicogenesis of *atomic nucleus*

An interesting example of pseudo-Classical neology, that is the influence of 'vulgar languages' on English, can be shown through the lexicogenesis of the term *tunnel diode* and the subsequent *unipolar resonant tunnelling diode* (Table 4).

TUNNEL	DIODE
1440: From the Old French *tonel* – tubular net 1839: n.[4] A subterranean passage; a roadway excavated underground especially a hill or mountain. Tunnel[ling + led + es]. From tunnel 1687 *v. trans* to catch (partridges) with a tonel 1856: v.t. [2c] To excavate, as a tunnel; To make (one's way) by boring or excavating. 1928: 'Possibility for the transmission of a particle through a [ ] barrier which would otherwise be insurmountable'.	(From Electrode – 1834; Greek: Electric + odos (way) ≡ One of the poles of a galvanic battery (anode and cathode)) 1919: [Compound/contraction of *Di* + *electrode*] A thermionic valve of the simplest kind, with just two electrodes (anode and cathode). 1950's : Diode (semiconductor): A diode constructed from semi-conducting material.
1960s TUNNEL DIODE: Junction diode with such a thin depletion layer that electrons bypass the potential barrier.	
C. 1970S BACKWARD DIODES – UNI-TUNNEL DIODES	
c. 1980s Resonant tunnelling diode	
c. 1990s Unipolar resonant tunnelling diode Bipolar tunnelling diode	

Table 4: Lexicogenesis of *tunnel diode*

Scientists and technologists, it appears, borrow from other languages and other specialisms and make interesting compound words. Many of these neologisms survive and enter dictionaries, sometimes by way of neology collections. The compound terms sometimes survive intact, e.g. *nuclear physics* and *parallel computer*, and sometimes one of the constituents is omitted, for instance, the terms *program* and *nucleus*. Lexicogenesis may be used to monitor the growth of a new discipline.

Case Study II: Affixation in Dictionaries and Corpora

In order to look at the role of morphological processes in the creation of neologisms, we have examined three major dictionaries of linguistics, computing and physics (see Table 1 above), and looked at three

corpora developed at the University of Surrey (see Table 5). The first corpus comprises texts in linguistics, particularly dealing with papers on morphology and syntax, the second comprises texts on nuclear physics, especially nuclear structure physics, and the third is a corpus of electronic texts focused on a newly-emergent form of device called Tunnel Diode. Each corpus contains genre-varied texts including texts from learned journals, textbooks, doctoral dissertations, popular material and informal texts like announcements of conferences and courses in each of these specialisms. With regard to content each corpus (as defined by subject field) focused on a particular topic: the emergence of typological studies in linguistics (c.1950-1990); the discussion about the structure of nuclei during the pre-war years (1900-1945) and that of unbound nuclei (1970-90); and the emergence of quantum devices (c. 1970-1990) in electronics which are yet to be fabricated. Each corpus attempts to document the genesis of an idea through the lexicogenesis of the vocabulary particularly through the morphological processes of suffixation and prefixation. The composition of the corpus is shown in Table 5 below:

Corpus	**Size**	
	No. of Texts	**Total No. of words**
Linguistics	**68**	**688,733**
Nuclear Physics	**171**	**580,470**
Semi-conductor Electronics	**94**	**434,600**

Table 5: Specialist corpora used in this study for suffixation and prefixation in specialist texts.

Case Study IIa: Suffixation

Nominalisation hallmarks specialist language and this appears to be a very productive, if at times frivolous use of such an important linguistic device available to the writers of texts. Nominalisations are nouns 'derived' from the verb (or adjective). Halliday and Martin have remarked that verbs are regrammaticised in scientific discourse into nouns; something which happens on a 'massive scale' in this discourse for reconstructing the nature of experience as a whole: 'The elaborated register of scientific knowledge reconstructs as an edifice of thing' (1993: 15): [things] which can be observed and experimented with. For instance, scientists will take *'stable, behave, occur, develop, useful'* and regrammaticise them, through derivational affixes, into *stability, behaviour, occurrence, development, utility*.

Biber, Conrad and Repen have conducted a contrastive study of texts of different registers with a view to investigating the 'distribution and function of nominalization' (1998:59-65). They have looked at four common derivations of nouns: two from words that are in the verbal category, namely nominalised words ending in *-tion/-sion* and *-ment*; and two from the adjectival category, namely nominalised words ending in *-ness* and *-ity*. The authors have also studied plurals of nominalised words. Biber and colleagues have examined the Longman-Lancaster Corpus and the London-Lund Corpus of spoken British English. Both corpora were tagged and hence the authors could authoritatively talk about word classes like nouns, verbs and adjectives. Their principal finding was that proportions of nominalisations used in the formal-informative register, *academic prose*, were very different i.e. from the imaginative register, *fiction*: Preponderant in the former and rare in the latter.

One characteristic of special languages is the use of number classes. Scientists and engineers use the two-term contrast: singular denoting unity, and plural used to denote classes of objects and events, and types of processes. Witness, for example, the discussion of the hitherto hypothetical unifying *force* in particle force which manifests itself as one of the four *forces* - electromagnetic, nuclear, weak and gravitational; witness also the predicated universal *grammar* which forms the basis of the grammars of (all?) natural language grammars.

We have looked at the distribution of nominalisations in two of our corpora. However, since our texts are not tagged, it will be difficult for us to be as certain about our results as Biber and colleagues have been. Nevertheless, our results are in broad agreement with Biber and colleagues and confirm Halliday's remark about scientists always attempting to build an edifice of things by extensively nominalising verbs.

Our study has focused on the nominalised words that were derived from verbs. The percentage of nominalised words derived from adjectives was very low. The derivational suffixes we report on are -*tion/sion*, -*tions/sions*, and -*ment/ments*. There is a brief description of nominals derived from adjectives also based on the comparison of the nominals ending in -*ness(es).*

In nuclear physics, the most nominalised verbs include *react*, *calculate*, *interact*, *distribute*, and *radiate*. These verbs are seldom used and their nominalised forms, *reaction(s), calculation(s), interaction(s), distribution(s)* and *radiation(s)* tend to dominate the texts: only 5 instances of *react* were found in the nuclear physics corpus as compared to 971 instances of *reaction(s)*. The percentage use of the nominalised form is over 90% for the five nominalised words discussed above (See Table 6):

Token	Base Word	Nominalised Form	Base word + All suffixations	%age Base words	%age Nominalised Form	Comments
react	5	971	980	0.51%	99.08%	
calculate	76	932	1030	7.38%	90.49%	calculates (1)/ed/ing(21)
interact	21	728	808	2.60%	90.10%	
distribute	3	655	659	0.46%	99.39%	
transit	6	399	405	1.48%	98.52%	
radiate	3	397	401	0.75%	99.00%	
approximate	68	390	600	11.33%	65.00%	approximately (138)

Table 6: The Surrey Nuclear Physics Corpus: The distribution of the 7 most frequent nominalised words ending in *-tion/-sion* and their plurals together with the verb base from which the nominals were derived.

Our linguistics corpus comprises a large number of texts which discuss the morphology of languages as well as study of language. A number of texts that emphasise the role of lexis in the study of language, particularly the morphological. Hence nominalised terms like *agreement, inflection* and (case/gender) *assignment* have a high frequency. Most of the linguistics deals with the *'construction'* of grammar(s) and other artefacts: hence there are terms related to the *violation* of rules (of grammar), *assignment* of categories and features, and there is description of *relation(s)*. Table 7 shows some of the most frequently occurring nominalised verbs in our corpus; some contributing to over 90% usage of the nominalised form of the base token verb.

Token	Base Word	Nominalised Form	Base word + All Suffixations	%age Base words	%age Nominalised Form	Comments
agree	98	+ment 2280	2508	3.91%	90.91%	agrees (57), agreed/ing (73)
relate	33	+ ion 727	836	3.95%	86.96%	relates (41)
assign	84	ment 540	947	8.87%	57.02%	assigned/ing (285)
inflect	3	tion 351	398	0.75%	88.19%	
construct	37	306	391	9.46%	78.26%	constructed/ing (34)
violate	33	216	292	11.30%	73.97%	

Table 7: The Surrey Linguistic Corpus: The distribution of the 6 most frequent nominalised words ending in *–ment, -ion, -tion* and their plurals together with the verb base from which the nominals were derived.

Suffixation is a productive device used by scientists and the derived forms of verbs have a significant iconic value here. The nominalised terms become the seed of a whole array of compounds used not only to indicate the developments within a sub-discipline but also to create the edifice of concepts and artefacts of new (sub) discipline. From *react* we have had *reaction(s), reactor(s) and reactant(s)* and on to chemical/nuclear *reactions* and subsequently to chemical/nuclear *engineering*!

Case Study IIb: Prefixes

Now we take a look at prefixes, which Ayto suggests are the least resilient form of word formation.

When we look at the use of prefix words in our corpora, they form a substantial part of the lexical inventory. Some of these prefixes are used to form words in general language (words like *incomplete, non-trivial* and so on) but the others are used with nouns, adjectives and verbs with a specialist sense, for example, *non-central* as in *non-central forces* in physics, *coargument* and its plural *coarguments* and abbreviation *coarg.*, as well as *cophonology, transformation* (and its variants *transformational, transformations* and *transformed*) in linguistics, *multi-access* in computing, *anti-bonding* and *demultiplexing* in the tunnel diode literature, and *antiproton, inelastic, pseudovector* as found in nuclear physics.

Prifixed words are used to negate, reverse, indicate degree of size or orientation and attitude, show location, time and order or number related to an established concept within the discipline. For instance, *antibonding, depopulation, interband, non-local, overlayer, pseudopotential, renormalise, submicron, superlattice* and *undoped,* would not exist without concepts related to *bonds, populations* (of electrons), *bands* (in solids), *potentials* (which were not local*), layers* (of atoms*), potentials,* (the concept of) *normalisation, microscopic dimensions, lattices* and *doping* (of electronic material to change their characteristics). Similarly in nuclear physics we have *antikaons, antiprotons* to indicate the pre-existence of protons and neutrons, *hyperdeformation* to indicate degree of deformation. Linguists talk about *detransitivisation, extrametricality, nonmonotonocity, postnominal, subcategorisation* and *unaccusatives* only because they already have transitives, metrics, monotonocity, nominals, categories and accusatives respectively.

In some cases the prefixes are used to form *retronyms*: ' a modification of an existing term to distinguish it from a NEOLOGISM'. For example, *terrestrial television* is a retronym of *television* to distinguish it from *satellite television*. So in physics we first had *relativistic* speed and only then *non-relativistic* speed. Physicists have had *nuclei* which were stable systems, hence all nuclei were bound by implication. Only recently, have physicists been able to create *unbound* nuclei, and so we now have a term *bound nuclei* as well. The development of *quantum* mechanics led to the retro-definition of *classical* mechanics.

An initial analysis shows the use of terms which are essentially of the form prefix+term [+suffix]- found in three major specialist dictionaries, that of *Physics* (Issacs 1996), *Computing* (Illingworth 1996) and *Linguistics & Phonetics* (Crystal 1997) (described in Table 1 above). The most prominent prefixes in these dictionaries are the so-called 'neo-classical' prefixes (Quirk et al 1985). These neo-classical items account for up to a third of all prefixed terms in the Computing dictionary, a quarter in the Physics dictionary and about one-fifth in the Linguistics one. Quirk et al also have a 'miscellaneous' category for prefixes like *auto, extra, proto, self, semi* and *vice* amongst others which comprise the second largest category for prefixed terms. Closely following are prefixes that are used to indicate degree or size, or prefixes to indicate (polar) opposition through the use of the prefixes *non-* or *un-*. Prefixed terms on their own comprise about 5% of the total terms in the Physics and Computing dictionaries, but fewer (c.3%) in the Linguistics dictionary. This, perhaps, can be attributed to the dexterity with which linguists can manipulate language: it is easier for them, perhaps, when compared to the Physicists and Computing folk, to create a neologism by other means than prefixing. Table 8 shows an initially hand-counted analysis of the three dictionaries:

Prefix Category		Prefixes	Physics (1996)	Computing (1996)	Linguistics (1997)
1	Negative	Not (*non* + noun/adj/adv)	10	13	45
2	Degree or size	Extreme (*hyper* + adj/noun) More than (*super* + adj/noun/verb)	9 18	7 11	1 7
3	Orientation/attitude	Against (*pro/anti* + denominal-adj/noun)	8	2	1
4	Locative	Under (*sub* + adj/verb/noun)	6	20	11
5	Time/order	Back/again (p*re/re* + v/den. Noun-adj)	5	8	9
6	Miscellaneous	*auto, extra, neo, , self, semi, tele, vice*	19	57	23
7	Neoclassical items	Number prefixes (*bi, di, many, poly, uni, mono, multi*)	48	122	35
	Others		65	94	68
Total Prefixed Terms			188	334	200
Total Terms			c.4000	c.6000	c.6000

Table 8: Approximate distribution of prefixed terms in various specialist dictionaries. The 'others' category includes items related to the seven categories with lesser used prefixes.

An analysis of our three corpora (see Table 5 above), shows that there is a significant usage of prefixed terms. We distinguish between general language prefixed words (e.g. *unexpected, unresolved, indefinite, extravagant* etc) and prefixed terms. In our corpora we also found that not only does one find a prefixed term, but its suffixed variants as well. For instance, not only the prefixed term *sub-category* found in our corpus, but we also find *subcategories, subcategorization* (and *subcategorisation*), *subcategorizing* and *subcategorized*. (This example is not to start a *directionality* debate as to whether *subcategory* came first or *categorise*.) We have denoted the prefixed term as the lemma and treated others as the variants:

Corpus	Prefixed (+ variants)	Frequency
Linguistics	176	1363
Semi-conductor Electronics	145	3323
Nuclear Physics	130	4069

Table 9: Distribution of prefixed terms in our three corpora.

The distribution of the prefixes, according to the seven categories in Table 8, appears to be different in the dictionaries when compared to the specialist texts. For example, the authors of texts in our corpora do not use many of the so-called neo-Classical number prefixes. However, most of the prefixed terms, or at least, those used with some noticeable frequency (>5), exist in the dictionaries or have the root of the term itself in the dictionaries. Invariably, prefixed terms are used in their variant forms as well. The more frequently occuring prefixes in our nuclear physics corpus are used to indicate degree or size, for example *super heavy* as in *super* heavy elements – highly unstable large nuclei, and prefixes for negation as in *anti-proton* and *antiquark*. Prefixes like *pseudo* are regarded by Quirk et al as pejorative. However, in physics such a prefix is used to indicate that a certain concept is a hybrid of two already established concepts. For instance, a *pseudo-scalar* quantity is essentially a *scalar* that has magnitude, but partially exhibits properties of a vector, which has both magnitude and direction. Another interesting shift in prefixes in physics is from *hyper* to *super*: so we have *superdeformation* but less frequently *hyperdeformations*[1](see Table 10 for details)

[1] [(This may be a correction to the earlier usage of *hyper*, as in *hypernuclei*, when the reference was not made to size but the attachment of a *hyperon*, an elementary particle, to an otherwise stable nucleus)]

Prefixed Term	Dictionary	Freq.	*+suffixes to indicate number, tense and word class conversion*				
			+s	*+ed*	*+ing(s)*	*+tion(s)/al*	*ity/ly/ally/*
anti proton	√(L)	23	14				
anti quark	√(L)	7	3				
counter term		11	10				
hyperdeform				9		6	superdeformation (17)
hypernucleus	√	5	5				4 ally
inelastic	√	101					17 (17 ities)
nonlocal	√	21					
pseudoscalar	√	32	4				
renormalize	√	6	1	19		38	+able 2; ity 1
rescattering	√	16	1				
subshell	√	8	1				ivity 5
superconductor	√	3					
superheavy	√	32					
prefragment	√	5	1	42			
TOTAL		**164**	**16**	**28**	**42**	**44**	

Table 10: Key prefixed terms in the Surrey Nuclear Physics Corpus. The dictionary of specialist terms used here is *Dictionary of Physics* (1996). 'L' is used to indicate that a lemma relates to the term but not the prefixed form – e.g. the term *antiproton* does not exist in the dictionary but *proton* does.

The emergence of tunnel diodes is characterised by prefixes related to degree or size (*super*) and the locative *sub* as in *subband*, *substratum* and *substrata*. And we have pre-prefixed terms like *intersubband*. The key negative prefix is *un-* as in *undoped* materials. Terms in this subject are being borrowed from both physics and electronics. Note that many of the prefixed terms do not exist in the dictionaries of physics and electronics. (see Table 11).

Prefixed Term	*Dictionary*	*Freq.*	*+suffixes to indicate number, tense and word class conversion*					
			+s	*+ed*	*+ing(s)*	*+tion(s)/al*	*ity/ly*	*Other variants*
anticross	X		1		13 (2)			
decouple	√	1		6	5			
deform	√(L)	1	1			10		
depopulation	√(?L)	6		7				
discharge	√	5	2	3	11			
extrapolate	√(L)	2			2	6 (5)	(2)	
incoherent	√(L)	37						
interband	√(L)	50						*intersubband* (50)
non parabolic	√(?L)	5					(17)	
overlayer		31	21		1			
pseudopotential		4	2					
renormalize	Phy	1		2		14		*unnormalized* (2)
subband		117	108					
substrata	√	176	38					
superconduct(or)	√	10	12		47		5	
superlattice	√(L/Phys)	382	204					
undoped		86						

Table 11: Key prefixed terms in the Surrey Electronics Corpus. The dictionaries of specialist terms used here are the *Dictionary of Physics (1996)* and *Dictionary of Electronics (1998)*.

Our linguistics corpus shows the strong influence of computational linguistics. The prefixes 'pre-' and 'pro-' are amongst the most frequently encountered prefixes in our linguistic corpus. *Category*, and derivatives *categories*, *categorisation*, also exist in prefixed forms. Crystal's dictionary contains many

of these prefixed terms. (see Table 12).

Prefixed Term	*Dictionary*	*Freq.*	*+suffixes to indicate number, tense and category conversion*				
			+s	*+ed*	*+ing(s)*	*+tion(s)/al ity/ly*	*Abbreviations*
coargument	√(L)	31	6				11
coindex	√	1	1	45	4	14	
cophonology	√(L)	1	1				
coarticulation	√	1					
detransitive	X	0		5	2	3+4	
denominal	√	5					
deverbal	√	21					
extrametrical	√(L)	5				28	
infix	√	6	4	6	2	37	
non-configurational	√(L)	5				1	1
nonprototypical	√(L)	24					
postpose	√	2		6	8	9(7)(1)	
prefix	√	98	38	24	2	11	(preprefix 1+2)
prepoposition	√	93	71			(227)	
pronominal	√	209	45	1		9+7	
prototypical		24				3	
subcategorize	√	2+4	5	6+4		25(3)	9
subgender	√(L)	25	49				
subsegment	√(L)	33	26			(49)	72
superclass	V(L)	10	5				
transformation	√	19	36			(65)(1)	
unaccusative	√	8	4			7	
ungrammatical	√(L)	46				22	

Table 12: Key prefixed terms in the Surrey Linguistics Corpus. The dictionary of specialist terms used here is *Dictionary of Linguistics and Phonetics* (1997).

Our three corpora show that affixed words form a significant percentage of each of the corpora, generally in excess of 1% when we include prefix forms of both general and special language words. This is a significant proportion. Furthermore, it appears that, in subjects which have been well established like nuclear physics, one does not find as many prefixed words, whereas for the emerging subjects, where debate is quite heated, there are more prefixed words, like in linguistics, and the more recently emerging subjects have even more prefixed words like in tunnel diodes. One can argue that, even if they are briefly lived, prefixes per form a very valuable service, and the suffixes help scientists to create an edifice of concepts to explain natural and human phenomena.

Afterword

Neologisms are an important stock in trade of scientists and technologists and contribute to language growth and language change. Amongst the important class of neologisms are the extant words, words in the natural language of scientists and engineers and words in prestige/classical languages, which are relexicalised. After relexicalisation, the terms are affixed to form more neologisms and nonce formations. Juxtapositional neology and integral borrowings not only bring new words into a language but occasionally bring in new syntactic and morphological structures as well. Our study of specialist dictionaries and specialist corpora shows that both suffixation and prefixation are used in specialist texts. Both are used in creating the edifice of a science by the scientists using plurals and changing words classes, for example, by nominalization. The prefixes are used less frequently, largely to show contrast. Perhaps more importantly, prefixed terms are used for indicating concepts and devices that are not quite in the mainstream of scientific and technological thought. The use of negatives, intensifiers for degree/size and locatives shows this process at work. Prefixes may be shor-lived but act as important place holders whilst scientists resolve contradictions and discover new concepts and artefacts.

The words of the bard, William Shakespeare (76th sonnet), are perhaps the best summary for describing neological activities and challenges: (First cited by Otto Jespersen)

> So all my best is dressing old words new/ Spending again what is already spent:
> For as the sun is daily new and old/ So is my love still telling what is told

References

Dictionaries

Ayto, John. (1999). *Twentieth Century Words*. Oxford: Oxford University Press.

Barnhart, Clarence, L., Steinmetz, Sol., and Barnhart, K. Robert. (1980). *The Second Barnhart Dictionary of New English*. New York: Barnhart Books.

Crystal, David. (1997) *A Dictionary of Linguistics and Phonetics (4th Edition).* Oxford and Cambridge (Mass., USA): Blackwell Publishers Ltd.

Green, Johnathan. (1991). *Bloomsbury Neologisms: New Words since 1960*. London: Bloomsbury Pub. Ltd.

Hartmann, Reinhardt and James, Gregory. (1998). *Dictionary of Lexicography*. London: Routledge.

Isaacs, Alan. (1996). *A Dictionary of Physics* (3rd Edition). Oxford and New York: Oxford Univ. Press.

Illingworth, Valerie. (1998). *Dictionary of Electronics.* London: Penguin Books.

Illingworth, Valerie. (1996). *Oxford Dictionary of Computing.* Oxford: Oxford Univ. Press.

Papers and Books

Aitchison, J (1991). *Language Change: Decay or Evolution*. Cambridge Univ. Press.

Ayto, John (1995). 'Lexical life expectancy – a prognostic guide'. In (Ed.) Jan Svartvik *Words: Proceedings of an International Symposium, Lund, 25-26 August 1996*. Stockholm: Kungl. Vitterhets Historie och Antikvitets Akademien. pp181-188.

Bailey, Richard, W. (1987) (Ed.) *Dictionaries of English: Prospects for the Record of our Language*. Cambridge: Cambridge University Press.

Gell-Mann, Murray (1997). 'Three Quarks for Muster Mark'. Interview transcript in (Eds.) Lewis Wolpert and Richards, Alison. *Passionate Minds: The Inner World of Scientists*. Oxford: Oxford Univ. Press. pp.159-165.

Halliday, Michael A. K., and Martin, John R. (1993). *Writing Science: Literary and Discursive Power*. London & Washington DC: The Falmer Press.

Hayes, Donald P. (1992). 'The growing inaccessibility of science'. *Nature*. Vol. 356. pp. 739-740

Jespersen, Otto. (1938/1962). *Growth and Structure of the English Language*. (9th edition). Oxford: Basil Blackwell.

Lehrer, Adrienne (1996). 'Why neologisms are important to study'. *Lexicology vol. 2/1*. pp. 63-73.

McCulloch, Warren, S. and Pitts, Walter H. (1990). 'A Logical Calculus of the Ideas Immanent in Nervous Activity'. In (Ed.) Margaret Boden. pp. 22-39. (Originally published in 1943).

McMahon, April S. (1994). *Understanding Language Change.* Cambridge and New York: Cambridge Univ. Press.

Picone, Michael, D. (1996). *'Anglicisms, Neologisms and Dynamic French'* John Benjamin Publishing Company

Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey and Svartvik, Jan. (1985). *A Comprehensive Grammar of the English Language.* London and New York: Longman.

Von Neumann, John L. (1958). *The Computer and the Brain*. New Haven: Yale Univ. Press.

Weiner, E S C. (1987). 'The New Oxford English Dictionary: Progress and Prospects'. In (Ed.) Richard W Bailey. pp. 30-48.