Productivity and Diachronic Evolution of Adjectival and Participial Compound Pre-modifiers in English for Specific Purposes

Alexandra Mestivier-Volanschi

Abstract This article studies the ways in which adjectival and participial compound pre-modifiers are used in scientific texts in the field of biology. We do not attempt to strictly classify these pre-modifiers within the class of adjectival compounds or dispute the fluctuating terminology in this field of morpho-syntactic analysis. Instead, we aim to provide detailed and reliable statistical data on the productivity of this phenomenon which we claim is an undeniable characteristic of at least two genres of the scientific register: research articles and monographs. Using data derived from several specialised and reference corpora, scientific journal archives and monograph collections we show that this type of construction is significantly more frequent in English for Specific Purposes (ESP), at least in the fields of Biology and Earth Sciences, than in English for General Purposes (EGP) and that it has become so especially over the last half-century. The article is structured as follows: Section 1 reviews a number of aspects related to Adjectival and Prepositional Compound Pre-modifiers (APCPs) discussed in the literature. While justifying terminological and orthographic choices, the article discusses the types of morpho-syntactic process which are at play in the production of APCPs as well as the role these compounds play in building information structure. Section 2 presents the resources used in the statistical study of the phenomenon while section 3 measures the productivity of APCPs using several qualitative and quantitative measures both synchronically (in specialised vs. reference corpora) and diachronically (over the last 100 years).

Keywords Compound pre-modifiers, adjectival compound pre-modifiers, participial compound pre-modifiers, ESP, morphological productivity, diachronic evolution

1 Adjectival and participial compound pre-modifiers

Adjectival and participial compound pre-modifiers (APCPs) share a number of formal similarities: they are headed by adjectives or (apparent) participles, past or present, while their non-heads are arguments, adjuncts or modifiers of the heads. Within this set of compounds, all equivalent to full or reduced relative clauses, we identified four classes:

Class 1: **Participial compound pre-modifiers (PCP)** built on past participles, illustrated by examples in a–d:

a) *toluene-fixed* cells (cells fixed with 1% toluene), *head-mounted* camera (mounted on the head),
b) *in vitro-determined* kinetic properties, *positively-charged* amino-acids,
c) *virus-infected* cells, *mitogen-activated* protein kinase,
d) *food-restricted* birds, *calorie-restricted* diet, *progeny-tested* bulls, *cohesin-enriched* domains.

These PCPs are built on a past participle attached to a) prepositional objects with various roles (agentive, instrumental, locative), b) modifying adverbials or c) passive subjects. The verb-direct object relationship within the PCP is unlikely, since the object role is typically played by the noun that the PCP modifies. When the non-head element seems to be a direct object of the past participle, the structure may actually be interpreted as being derived from a nominal construction, by adding the participial suffix *-ed* (d). This is confirmed by the occurrence of the alternative forms *food restriction*, *calorie restriction*, *progeny-testing.* Alternatively, Roeper and Siegel (1978: 234 f.) suggest that constructions such as *time-controlled*, *surface-sealed*, *tongue-tied*, can be paraphrased as a passive with a prepositional phrase: *controlled in time*, *sealed at the surface*, *tied by the tongue.* We found similar paraphrases for PCPs such as *calorie-restricted*, *food-restricted*, *progeny-tested* in the specialised corpora we compiled for the study: *restricted in calories*, *restricted on food*, *tested for/with respect to progeny.*

Class 2: **Adjectival compound pre-modifiers (ACPs)** are built on a specifying adjective linked to one of its sub-categorised prepositional objects (a):

a) *replication-competent* virus (competent for), *electron-dense* aggregates (dense in), *drug-susceptible* isolates (susceptible to), *glycine-rich* regions (rich in), *allele-specific* primers (specific for), *serum-free* medium (free of), etc.

Within the class of ACPs, a sub-class of compounds is headed by adjectives derived from verbs and having inherited the verb's argument structure (b):

b) *time-dependent* analysis, *cold-responsive* genes, *tissue-protective* effect, *virus-resistant* transgenic plants.

Class 3: nominal base + *-ed* constructions, also called **possessive compounds** (Bauer 2010: 169), **derivational compounds** (Adams 1973: 99–101) or **extended bahuvrihi compounds** (Kastovsky 1992: 389), illustrated by examples a–c:

a) *single-celled* eukaryote, *thick-walled* micropipettes,
b) *rod-shaped* bacteria,
c) *two-tailed* t-test, *three-tiered* structure.

Compounds in this class are only apparently headed by a past participle; they are actually built by derivation (hence the term "derivational compounds"), by adding the participial suffix *-ed* to a) an adjective-noun, b) noun-noun, c) quantifier-noun combination. These constructions may be interpreted as attributing a property (designated by the compound) to the modified noun: a straight-winged male is a male which *has* straight wings (hence the term "possessive/possessional compounds").

Class 4: **Participial Compound Pre-modifiers** based on a present participle attached to its a) direct object or b) adjunct:

a) *colony-forming* units, *quorum-sensing* regulators, *receptor-binding* activity,
b) *high-scoring* sequences, *fine-mapping* studies, *long-acting* drugs, *fast-growing* cells.

The noun modified by the PCP is the verb's subject, while the non-head element within the PCP is most commonly the direct object of a transitive verb.

In the linguistics literature, these types of compounds have been considered to be more marginal than nominal compounds and have therefore been given much less attention.

According to the statistical data which we gathered on specialised and reference corpora, it would seem that, while this may be the case in EGP, the situation is completely different with the specialised languages studied. This can be seen with some initial data: while the ratio of APCP tokens to nominal compounds was found to be 17.7 to 51.7 in the CoCA corpus, in the PLoS corpus the same ratio was 34.3 to 51.5 (both corpora are described in section 2). And when types rather than tokens are considered, this ratio was 42.0 to 44.9.

The present study provides further statistical data on the productivity of these classes of compound pre-modifiers in present-day ESP, and their diachronic evolution over the past century. But before we move on to the presentation of the corpora used and of the statistical data obtained, we would like to clarify a number of terminological and orthographic choices made in the study and briefly review several theoretical points related to the roles APCPs play in scientific discourse.

1.1 Terminological choices

The area of compounding in English is characterised by considerable terminological variability. Various terms have been used for the units in classes 1–4 set out above either separately or as a group, depending on the aspect for which they were considered, or on the theoretical framework of the linguist discussing the phenomenon. While several studies (Marchand 1969, Adams 2001, Conti 2006) classified all four classes under the heading "compound adjectives", different classifications either include or exclude certain types of compounds.

Participial Compound Pre-Modifiers (PCPs): Marchand (1969) calls them "verbal nexus adjectives in *-ing* and *-ed*" distinguishing between first and second participles, Adams (2001) distinguishes between *-ing* and *-ed* compound adjectives, Conti (2006) uses the term "participial compounds".

Other authors may include or exclude PCPs in a class of compounds called "verbal or synthetic compounds" (*truck-driver*: noun-noun compounds having a de-verbal element as a base and an argument of the verb in the left-hand element), which are traditionally opposed to another class of noun-noun compounds, called "primary or root compounds" (*coffee table*). While some investigators limit the term "synthetic compounds" to those derived from action nominals and subject nominalisations (*train driving*, *train driver*), other scholars (Roeper/Siegel 1978) discuss only compounds whose second element ends with one of the suffixes *-er*, *-ing* and *-ed*, while others still (Lieber 2004) specifically include in this category compounds formed with other nominalising suffixes like *-ion, -ment, -ure, -al, -ance, -ation.*

The area of "synthetic compounding" in English has been rightly described as a "descriptive and terminological nightmare […] totally unhelpful […] [leaving] subsequent scholars in a state of perpetual confusion in trying to interpret the competing claims made about synthetic compounding" (Bauer/Renouf 2001: 117). We decided not to take any position in the discussion on where the boundaries of synthetic compounding lie and avoided using this terminology.

Adjectival Compound Pre-modifiers (ACPs): The term proposed by Conti (2006: 74), "restrictive/specifying compounds", is an appropriate name for the class: with these compounds, the adjectival head is associated to a prepositional object which specifies one aspect of the adjective. In a *replication-competent virus*, the non-head member of the ACP specifies the

virus' domain of competence. However, for a symmetrical presentation with PCPs, we decided to systematically use the term **adjectival pre-modifier compound**, which focuses on the configurational rather than the semantic aspect of these compounds.

1.2 Orthographic choices

We freely admit that, since hyphenated forms are the only ones which can be found automatically (without considerable data in the way of heuristics – i. e. linguistic co-occurrence signals), this was our primary reason for focusing on them alone. In practice, non-hyphenated APCPs formed by argument linking (a) may not be easily automatically distinguished from reduced relative clauses (b) or plain subject-verb couples in finite clauses (c).

a) A putative relationship between hyperglycemic hormone degradation and **stress induced** protease activity in the hemolymph of the American cockroach.
b) These results demonstrate that **stress induced** by handling disrupts the normal capacity to osmoregulate in P. decipiens. *(stress which is induced by …)*
c) [Capture **stress**] **induced** a 25 % reduction in spleen haemoglobin concentration …

But in fact, there are good linguistic reasons for favouring hyphenated forms. First, an orthographic principle, even though inconsistently applied, which stipulates that compounds should be spelled as two words when used independently and as one hyphenated word when used as pre-modifiers. Given this principle, it is natural to assume that hyphenated forms of APCPs are much more frequent than their non-hyphenated equivalent. To take a single example, the JEB corpus (cf. section 2) contains 350 occurrences of *stress-induced* for 99 instances of *stress induced*, 67 of which are not instances formed by argument-linking. This leads to the second reason for which we chose to focus on hyphenated compounds. Hyphenation is a mark of argument linking; it is the preferred orthographic form as it may often play a disambiguating role in long, complex nominal constructions, rendering them more readable. When processing an article title such as *Endogenous opiate involvement in acute and chronic stress-induced changes in plasma LH concentration in the male rat*, the hyphen signals that a PCP is being used, immediately blocking the alternative reading: *[chronic stress] induced [changes]*.

1.3 APCPs and the textual information structure

Several reasons for the productivity of APCP-building mechanisms have been invoked in the literature.

The information-packing function. Biber (1988: 105) pointed out that attributive adjectives are "a more integrated form of nominal elaboration than predicative adjectives or relative clauses, since they pack information into fewer words and structures". Replacing a post-modifying relative clause by a semantically equivalent compound pre-modifier taking up much less space is an efficient means of compacting information. This has been shown to be the case for the written press where pre-modifying compounds "maximise a text's newsworthiness by packing the maximum content into the minimum number of words" (Ljung 2000: 208). We will show that this form of information compression is a convention adopted by scientific articles as well.

This "iconising function" (Opizzi 2006: 91) of compounds in general has been associated with an attention-grabbing quality provided by their conceptual richness and syntactic compaction and with a potential for concept formation (labelling or referential function): "The combination of two or more terms produces not only a mere union of existing concepts but the premodification process often gives rise to a new concept that alters their nature; adding new meaning and uses." (Gotti 2008: 77)

Building information structure. Moving content modifiers from post-nominal to a pre-modifying position is a means of presenting the information they convey as *given* or *presupposed* in a rhetorical effect of *backgrounding* (Halliday/Martin 1993: 60), either because the new information was presented previously in the text (in this case compounds have a cohesive function in the text) or because it is assumed to be a part of the specialised knowledge shared by a scientific community (this may be assumed to be the case with lexicalised, frequently used compounds). In the former case, compounds are created ad-hoc, as the text is produced; they are prepared in the preceding text. In section 3.2 we provide the proportion of APCPs which appear only once in the PLoS corpus; hapaxes are an indication that this mechanism is productive, i. e. it is used whenever the author needs to refer to previously presented information in a concise manner.

Adopting a terminology proposed by Francis (1994) in her discussion of lexical cohesion, Ormrod (2004) analyses the construction of complex nominal groups in scientific articles in terms of *advance* and *retrospective constructions*. In advance constructions "a nominal group […] is presented without any preparation during the preceding discourse", while retrospective constructions are "created as the text unfolds, re-using terms already presented" (Ormrod 2004: 54). Confirming Ormrod's observations we found that advance constructions are typically found in article titles, sub-headings and abstracts, and are subsequently rendered explicit in the introduction of the article (example 1) while retrospective constructions are created in subsequent article sections (Results, Materials and Methods, Discussion) (example 2).

Example 1. Advance construction[1]: the compound is announced in the introduction, rendered explicit in the Results section, then used as part of a coherence chain.

Line 12 (Introduction): This polypeptide […] acts as a major target of the erythrocyte **invasion-inhibitory** Ab response in individuals immune to Plasmodium falciparum malaria .

Line 40 (Results): mAb 12.8 and 12.10 have been shown to **inhibit erythrocyte invasion** in vitro by P. falciparum merozoites […] We therefore engineered the most promising scFvs into fully human antibodies since the presence of the Fc may potentiate **inhibition of erythrocyte invasion**, a likely prediction given that Fab and F(ab,) 2 fragments of mAb 12.10 do not retain **invasion-inhibitory** properties .

Line 60 (Discussion): […] with similar affinities to mouse mAbs, 12.10 and 12.8, already known to **inhibit erythrocyte invasion** in vitro.

[1] McIntosh RS et al. (2007): "The Importance of Human FcγRI in Mediating Protection to Malaria." *PLoS Pathogens* 3.5: e72. doi:10.1371/journal.ppat.0030072. <http://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.0030072> (March 2015).

Line 62 (Discussion): However, **inhibition of invasion** assays are done with heat-inactivated serum or in the presence of Albumax in the place of serum, suggesting that complement does not play a role, at least in vitro.

Line 93 (Materials and Methods): Antigen binding and **invasion–inhibitory** assays

Example 2. Retrospective construction[2]: the compound is prepared for in the introduction, then used in alternation with equivalent phrasal constructions.

Line 2 (Abstract): the computer algorithm **Ahab** […] **predicts** many novel modules within the network and genome-wide […] Moreover, we demonstrate for the entire set of known and newly validated modules that **Ahab's prediction** of binding sites correlates well with the expression patterns produced by the modules […] Finally, by comparing **Ahab predictions** with different categories of transcription factor input, we confirm the regulatory structure of the segmentation gene network.

Line 17 (Introduction): (B) **Ahab-predicted modules** in the control regions of segmentation genes were classified based on their composition into pair-rule driven [...]

Line 26 (Introduction): We used **Ahab** for a genome-wide **prediction** of segmentation gene modules with maternal and gap input […]

Line 28 (Introduction): Furthermore, we systematically analyze **Ahab's prediction** of binding site composition for all experimentally validated modules. […] Finally, we explore **Ahab's predictive** ability when binding site information is less well defined, as is the case with the pair-rule factors.

Line 40 (Results): Figure 1: **Ahab Predictions** and Recovery of Known Modules

Line 48 (Results): Under these conditions, **Ahab predicts** 52 modules within the genomic region of the 29 genes of interest, an average of about two modules per gene.

Line 51 (Results): Position of modules **predicted by the Ahab** mg run relative to the transcription start site of the cognate loci …

Line 56 (Discussion): In this study we have demonstrated that the **Ahab algorithm** can be used successfully for two purposes: the **prediction** of novel segmentation modules within genomic sequence and the **prediction** of module binding site composition.

Line 88 (Materials and Methods): **Ahab**, in its **prediction** of binding sites, fits all factors simultaneously. To gauge whether **Ahab can be used as a predictor** of module composition, we examined what fraction of known binding sites the algorithm recovers.

Line 95 (Materials and Methods): To assess the conservation of known and **Ahab-predicted modules**, we aligned D. melanogaster and D. pseudoobscura genomic sequence.

Line 99 *(Materials and Methods)*: To associate **predictions from different Ahab runs**, each run was processed and the highest point on the free energy plot within an interval of the window size was marked as a "peak."

Line 138 (Supporting Information): Sequence Information for **Ahab-Predicted** Modules in the Control Regions of 48 Segmentation Genes.

[2] Schroeder MD, Pearce M, Fak J, Fan H, Unnerstall U, et al. (2004): "Transcriptional Control in the Segmentation Gene Network of Drosophila." *PLoS Biology 2.9*, e271. doi:10.1371/journal.pbio.0020271 <http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0020271> (March 2015).

1.4 A form of syntactic re-packaging

APCPs are obviously paraphrases of a semantically related relative clause; some authors go as far as to use the term *premodifying relative clause* (Cortés/Verdejo 2006) instead of compound, which places them closer to a syntactic rather than a lexical phenomenon. Precisely where the border between compounding in general as a lexical or syntactic process lies is a matter of controversy. Several approaches analysed compounds as diachronically derived from semantically corresponding syntactic constructions (cf. the *condensation hypothesis,* Wälchli 2005: 246). Transformationalist approaches claimed that compounds are condensed sentences, derived through progressive transformations: "No surface structure classification of compound adjectives is possible without a preceding (implicit or explicit) analysis which is based on their underlying structure" (Meys 1975: 80). The mechanisms by which arguments or adjuncts of a verb become attached to a compound base have been modelled as lexical transformations (Roeper/Siegel 1978: "the First Sister Principle"), or according to interpreting principles based on argument structure (Lieber 1983: "Argument Linking Principle", or Selkirk 1982: "First Order Projection Condition"): lexical items (stems and affixes) have lexical argument structures which must be satisfied when inserted into a syntactic and lexical tree the same way they are satisfied when inserted in a syntactic tree of a sentence (Lieber 1983: 251).

Kastovsky (1995) places the functions of morphological processes along a continuum ranging from pure lexical coinages to what may be called forms of "syntactic repackaging". While the prototypical function of the former is to provide names for "nameworthy segments of extralinguistic reality" (Kastovsky 1995: 157), the second contributes to text cohesion and condensation (cf. section 1.3). An example of the former would be coining by derivation using a suffix such as *-ness*, *-ment*, *-ity*, or nominal compounding, while APCPs would be typical examples of the latter. Ljung (2000: 207) notes that the position along this cline from lexical term-creation to syntactic repackaging may be associated with differences in productivity: "naming is basically an idiosyncratic process, while syntactic repackaging is fully productive". It may easily be verified by a corpus query that few neologisms are created using the suffixes *-ness* or *-ment*, while the number of new APCPs is overwhelming. Section 3 hereafter offers several measures of the productivity of this method of syntactic reduction.

A further argument in favour of the idea that APCPs are more closely related to syntax than to the lexicon comes from their low degree of lexicalisation, as evidenced by their poor representation in present-day dictionaries. We analysed the few occurrences of APCPs in a number of specialised dictionaries (listed in Annex A) and identified a number of reasons for their presence among dictionary entries:

a) They may be part of a lexicalised term, usually accompanied by an acronym: *Artemisinin-based combination therapy* (ACT), *antigen-binding cell* (ABC), *Melanoma-associated antigens* (MAGE),
b) or part of an outdated term: *calcium-dependent regulator protein* (CDRP): early name for calmodulin, *antibody-producing cell*: equivalent to plasmacyte and plasma cell.
c) They may instantiate or render explicit a term headed by a lexeme denoting a generic category (*factor*, *complex*) or class of biological entities (*proteins*, *antigens*), while adding encyclopedic information to their terminological definitions, such as information on their function (all examples come from the Dictionary of Cell and Molecular Biology):

- *apoptosis-inducing factor*: A flavoprotein, 57kDa, which shares homology with the bacterial oxidoreductases
- *anaphase-promoting complex*: An unusually complicated ubiquitin ligase, composed of 13 core subunits and either of two loosely associated co-activators
- *actin-binding proteins:* A diverse group of proteins that bind to actin and that may stabilize F-actin filaments, nucleate filament formation, cross-link filaments, lead to bundle formation, etc.
- *melanoma-associated antigens*, a superfamily of proteins associated with tumours and part of the larger 'cancer/testis antigen' family.

d) It may be that, as pointed out by Gotti (2008), pre-modifying compounds play a labelling, referential function. In this case, the headwords could not be interpreted by virtue of a general rule schema: knowledge that a *colony-forming unit* is a unit which forms colonies does not suffice to understand the concept and the explanation provided here is essential: "Irradiated mice can have their immune systems reconstituted by the injection of bone marrow cells from a non-irradiated animal. The injected cells form colonies in the spleen, each colony representing the progeny of a pluripotent stem cell." (Dictionary of Cell and Molecular Biology).

As shown in section 3.1 hereafter, most APCP heads, both restrictive adjectives and participles, are very frequently used, combined with a vast range of arguments. It has been suggested (Ljung 2000) that APCP heads – be they participles or adjectives – may be in an intermediate stage of a process of grammaticalisation from usage as full verb or adjective in a relative clause to a suffix status, via their use in pre-modifying compounds. APCPs heads such as the adjectives *-free*, *-poor*, *-conscious*, *-friendly*, *-rich*, or participles like *-based*, are particularly productive in present-day English and may be considered as semi-suffixes.

Grammaticalisation may be conceptualised as a series of layers or states through which languages pass in the search for a more economic end efficient communication. Among other instances of grammaticalisation, decategorialisation is the evolution along a grammatical "path or cline of structural properties, from a morphologically heavier unit to one that is lighter" (Hopper/Traugott 2003: 106). Decategorialisation is a tendency for nouns, verbs or adjectives to lose morphological and syntactic properties that would identify them as a full member of a major category and acquire new uses as minor categories or as affixes. An example of this mechanism would be the evolution of the noun *while* to the conjunction *while*, losing grammatical features which identified it as a noun: the ability to have articles or quantifiers, being modified by adjectives, being referred to by an anaphoric pronoun, etc. Similarly, when grammaticalised as PCP heads, verbs lose the ability to show variation in tense, aspect, modality. Once a grammaticalisation pattern is established, new compounds are formed without any re-thinking of the origin of the construction. Following Quirk et al. (1985: 1568), who contend that compounding may be viewed in fact as "prefixation with open-class items", Hart (1994: 140) goes as far as to state that there is no fundamental difference between composition and derivation since many modern derivative suffixes were originally full or free second elements of compounds which lost their full meaning and taken on suffix function through decategorialisation: *-dom*, *-hood*, *-ship*, *-ly*, *-some* are considered today to be derivational suffixes. We may anticipate that this will become the case with ACP heads such as *-specific*, *-dense*, or PCP heads such as *-derived* or *-driven*, at least in the specialised languages we studied.

2 The corpora/resources used

In order to undertake a synchronic/diachronic statistical analysis of APCPs we consulted several reference and specialised corpora, as well as the archives of two scientific journals published over 8–9 decades.

Specialised corpora. In order to study the characteristics of Scientific English, we used two specialised corpora compiled for the illustration of two scientific domains.

1. *The PLoS corpus.* This 17 million word specialised corpus which we compiled is composed of research articles from 4 of the Public Library of Science publications: PLoS Biology (2003–2009: 5.2 million words), PLoS Genetics (2005–2009: 4.9 million words), PLoS Computational Biology (2005–2009: 3.5 million words) and PLoS Pathogens (2005–2009: 3.7 million words). PLoS journals articles are published under the Creative Commons Attribution License.[3]
2. *STEP corpus.* The STEP corpus is a 25 million word specialised corpus compiled by students in terminology and specialised translation from the University Paris Diderot, over several years, for a translation/terminological analysis academic project. It is constituted of research articles, PhD theses, manuals gathered by students in order to illustrate several disciplines of Earth and Planetary Sciences: volcanology, climatology, seismology, orogeny, etc.

Reference corpora. Two reference corpora were used in this study in order to contrast the usage of APCPs in EGP to the two varieties of Scientific English used in the bio-medical field and earth and planetary sciences.

1. *The BNC*. The British National Corpus is a 112 million word collection of samples of written and spoken language meant to represent a wide cross-section of British English from the later half of the 20th century.
2. *The CoCA*. The Corpus of Contemporary American English is a large and balanced 450 million word reference corpus for American English (1990–2012). The corpus is evenly divided between five genres of spoken, fiction, popular magazines, newspapers and academic journals.

For the purposes of the present study the scientific/academic parts in both corpora were excluded.

Scientific journal archives. In order to outline the diachronic evolution of APCPs, we consulted the archives of two scientific journals spanning several decades:

1. *Microbiology and Molecular Biology Reviews* (MMBR: published since 1937) is a scientific journal focusing on the "latest" developments in microbiology as well as related fields such as immunology and molecular and cellular biology. Its review articles explore the significance and the interrelationships of the latest discoveries on bacteria, viruses, parasites, fungi and other eukaryotes. Review sizes increased from an average of 18000 words in the 1930s to 24000 words in the 1990s. A corpus of sample articles from MMBR was built to study the evolution of APCPs over time.
2. *The Journal of Experimental Biology* (JEB: published since 1923) is a leading journal in comparative animal physiology publishing papers on the form and function of living or-

[3] The corpus may therefore be rendered accessible (please contact the author).

> ganisms at all levels of biological organisation (from the molecular and sub-cellular to the integrated whole animal). Under the JEB Open Access model, all users have unrestricted rights to re-use Open Access content for research purposes. We therefore built a sample corpus based on JEB articles to quantify the evolution of the APCP-building mechanism over time. Articles published in the JEB also increased from an average of 5600 words in the 1920s to 7600 words in the 1990s.[4]

Mini specialised corpus of monographs from the beginning of the 19th century. In order to obtain further evidence on the diachronic evolution of APCPs, we constituted a mini-corpus (1.5 million words) of journal articles and books in various fields of biology (microbiology, cell biology, biochemistry, bacteriology, genetics) dating from the 1860s to the 1910s. The legal copyright term of these documents has obviously expired so they were scanned by Google and rendered accessible via the OpenLibrary. A list of these documents is provided in Annex B.

3 Method and findings: statistical data on the productivity of APCPs

3.1 Qualitative and quantitative measures of morphological productivity

Before presenting the statistical data collected from our corpora, we briefly review here the statistical measures considered the most appropriate for quantifying productivity.

The definition of productivity we adopt here is the one provided by Bauer (2001: 211):

> "Productivity" deals with the number of new words that can be coined using a particular morphological process, and is ambiguous between the sense "availability" and the sense 'profitability'. The availability of a morphological process is its potential for repetitive rule-governed morphological coining […] [it] is determined by the language system, and any process is available or unavailable, with no middle ground […] The profitability of a morphological process reflects the extent to which its availability is exploited in language use and may be subject unpredictably to extra-systemic factors.

The two aspects of productivity (the general potential to create new words and the degree to which this potential is actually exploited by speakers) were further discussed by Plag (2003, 2006) in terms of the best suited qualitative and quantitative approaches to productivity. The availability of a morphological process is generally tested by counting the number of attested types (i. e. different words) at a certain point in time, in an unabridged dictionary. The problem with this approach is that, for instance, many entries in a present-day dictionary were formed using the nominalising suffix *-ment*; however, the suffix may not be said to be productive (most forms were created between the 16th and the 19th century). Similarly, as shown is section 1.4 above, very few APCPs are listed in specialised dictionaries. As the availability of a morphological process, viewed as a qualitative aspect, is not uncontroversial and may be difficultly verified, most approaches to productivity have focused on quantitative measures of profitability. Baayen (1993) proposed a number of corpus-based measures of productivity which rely on the availability of large electronic corpora, most of which were adopted here:

- *Type-frequency.* Measuring the number of types, or different words formed using a given affix in a corpus. This measure was adapted here by computing the number of different forms using a given base.

[4] The corpus may therefore be rendered accessible (please contact the author).

- *Number of hapax legomena* (words that occur only once in the corpus). Productive processes are expected to show large numbers of low frequency examples and small numbers of high frequency examples, "with the former keeping the rule alive" (Plag 2006: 542). It is among hapaxes that the highest proportion of neologisms can be found, especially in large-size corpora.
- *Productivity in the narrow sense.* The measure is based on the ratio of the number of hapaxes built with an affix and the number of all tokens containing that affix. In the present study this measure was adapted by dividing the number of APCP hapaxes built on a base (adjective or participle) by the number of types (instead of tokens) that the base gives rise to. We have thus replaced the probability of encountering an unattested word with a base with the proportion of hapaxes among all types.

In addition to the quantitative measures proposed by Baayen, we computed three extra measures, to study the usage of APCPs either in synchrony or in diachrony (over a period of almost nine decades): the proportion of APCS tokens among all hyphenated compounds, the normalised frequency (frequency per-million words, PMW), and, to downplay the effect that differences in sample corpora sizes may play on frequency evaluation, a chi-square test, which is the obvious statistic test for comparing proportions, and interpreted it using p-values (significance test measuring the probability that the differences in normalised frequencies of APCPs in the 1930s and in the 2000s could be the due to chance).

3.2 Frequency and quality of APCPs in reference and specialised corpora

Frequency lists of hyphenated compounds were established for each of the corpora studied: CoCA, BNC, PLoS, STEP. For each corpus we selected the most frequent *n* hyphenated compounds for close scrutiny (*n* is specified for each corpus in Table 1 below[5]). We then identified the APCPs in these lists in order to compute their respective frequencies in each corpus. We distinguished between ACPs, PCPs (past participle and present-participle-based) and possessive compounds (types 1–4 in section 1 above). Their respective type-frequencies estimated from the sample selected for analysis are reported in table 1 and graphically represented in figure 1:

Tab. 1: Type-frequency (percentage) of the four classes of APCP

CORPUS	**CoCA**	**BNC**	**PLoS**	**STEP**
Number of analysed compounds (types)	1000	2000	1500	750
Type 1: -Ved (ex. *food-deprived*)	9.10 %	13.15 %	21.79 %	22.0 %
Type 2: -Adj (ex. *region-specific*)	1.20 %	0.35 %	11.50 %	4.8 %
Type 3: -Ned (ex. *real-valued*)	2.10 %	3.35 %	0.53 %	1.73 %
Type 4: -Ving (ex. *oxygen-sensing*)	4.20 %	4.5 %	8.16 %	10.4 %

[5] Differences in the number of most frequent hyphenated compounds stem from differences in the length of the the frequency lists in the four corpora. We selected the *n* most frequent compounds having more than 10 occurrences in all corpora.

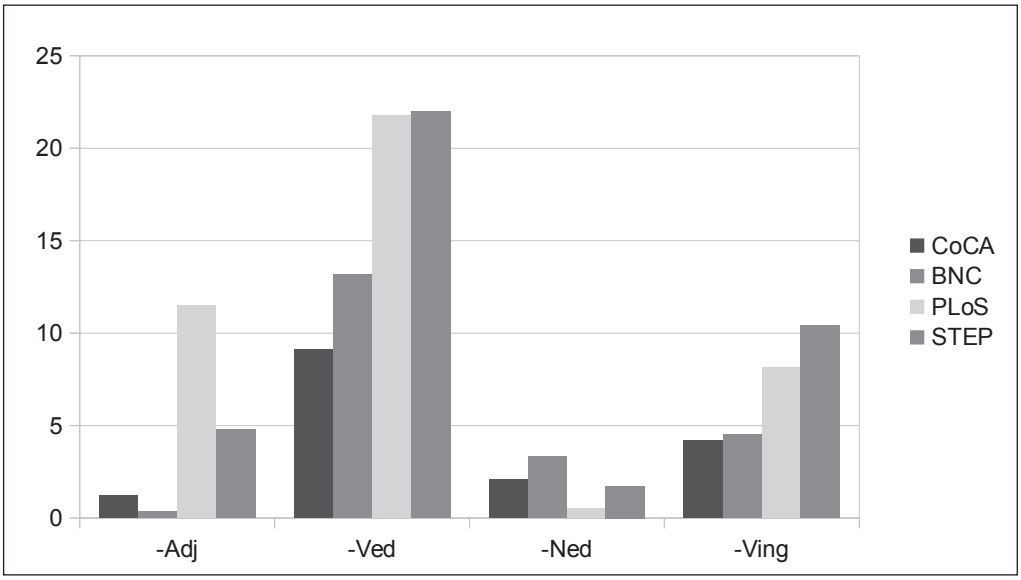

Fig. 1: Type frequency (percentage) of the four APCP classes

Several important observations can be made by analysing these data. Firstly, percentages of APCPs are considerably higher in specialised corpora than in reference corpora, for three of the four types studied, the only exception being the possessive compounds (*Type 3*), which is probably associated by scientists to literary/everyday language. Leaving this class aside, the use of APCPs clearly emerges as a characteristic of the scientific language used in experimental articles, at least in the fields of biology and earth and planetary sciences.

Secondly, in the specialised corpora, a number of adjectives and participles appear to be used recurrently for the creation of APCPs, forming "families" of compounds, characterised by a stable, predictable semantic relation between constituents (Conti 2006: 76). Some of these forms appear to be specific for each field of knowledge: compounds of the type *-parallel* (*fault-parallel, trench-parallel*), *-perpendicular* (*ridge-perpendicular, strike-perpendicular*), *-aligned* (*field-aligned, azimuthally-aligned*), *-dipping* (*north-dipping, inward-dipping*) are productive in the earth and planetary science research while compounds of the type *-resistant* (*drug-resistant, protease-resistant*), *-specific* (*tissue-specific, species-specific*), *-regulated* (*age-regulated, light-regulated*) or *-binding* (*ligand-binding, receptor-binding*) are abundantly used in APCPs in biology. Beside these forms, conceptually linked with each of the two fields of knowledge, a number of adjectives (*-dependent, -sensitive*), past participles (*-driven, -derived, -induced*) and present participles (*-causing, -forming*) appear to be equally productive in both fields. None of the adjectives or participles, past and present, in the two reference corpora were found to be particularly productive, with the possible exception of the form *-based* (10 compounds in the most frequent 1000 hyphenated compounds in the CoCA corpus and 10 compounds in the most frequent 2000 compounds in the BNC).

Before moving on to the diachronic study of the APCP-building mechanism, we selected a sample of ACP and PCP types (past participle alone), ranging from the most frequent to the least frequent ones. Table 2 lists their heads together with the number of different compounds

they give rise to (type frequency), their total frequency (token frequency), and the number and percentage of hapaxes in the PLoS corpus. As may be observed from the table, for most of these compound heads, hapaxes represent 50–60 % of the total number of types. As discussed in section 3.2, the percentage of hapaxes among compound types is a clear indicator that compound pre-modifier creation is a productive word-formation mechanism.

Tab. 2: Token and type frequency (including hapaxes) for a sample of APCPs

	-Adj (ACP)				**-Ved PCP**		
	Token Freq.	**Type Freq.**	**No. (and %) of Hapaxes**		**Token Freq.**	**Type Freq.**	**No. (and %) of Hapaxes**
-free	1341	430	109 **(25 %)**	*-bound*	697	129	65 **(50 %)**
-specific	7779	882	403 **(45 %)**	*-linked*	355	73	50 **(68 %)**
-dependent	3602	661	321 **(48 %)**	*-mediated*	2158	545	307 **(56 %)**
-sensitive	642	160	83 **(51 %)**	*-associated*	1883	445	236 **(53 %)**
-rich	729	146	66 **(45 %)**	*-driven*	406	145	85 **(58 %)**
-resistant	774	158	82 **(51 %)**	*-induced*	2370	628	356 **(56 %)**
-reactive	184	28	14 **(50 %)**	*-directed*	492	81	50 **(61 %)**
-tolerant	21	14	9 **(64 %)**	*-controlled*	200	54	32 **(59 %)**
-dense	110	13	8 **(61 %)**	*-coupled*	290	62	37 **(59 %)**
-competent	152	33	20 **(60 %)**	*-treated*	766	203	110 **(54 %)**
-susceptible	65	13	8 **(61 %)**	*-infected*	732	110	59 **(53 %)**

3.3 Diachronic study of -Adj and -Ved pre-modifier compounds

The mechanism for building APCPs was already available in Old English. Kastovsky (1992: 372–374) provides examples for each one of the classes we studied:

Type 1: *windfylled* 'blown down by the wind', *handgewriþen* 'hand-woven',
Type 2: *eagsyne* 'visible to the eye', *ellenrof* 'famed for strength',
Type 3: *feowerfotede* 'four-footed', cliferfete 'cloven-footed',
Type 4: *hunigflowende* 'flowing with honey', *rightfremmende* 'acting rightly'

With Middle English, the productivity of this pattern declined under the pressure of foreign influences; Latin borrowings replaced the formation of APCPs after the Norman Conquest: "The ME pattern of taking a ready-formed latinate loanword, rather than combining native items into a self-defining or transparent compound, reduced the presence of compounds in the language." (Hart 1994: 141)

But this type of compound has again become prevalent, especially in ESP, from the 19th century onward as we illustrate here for the field of biology; the process is probably associated with the growth of science, technology and industry, which rendered necessary the creation of

more complex pre-modifiers for newly discovered products, procedures, techniques, scientific facts, where simple adjectives were no longer sufficiently efficient and precise. The following sections track the variations in the frequency of APCPs based on the sample of heads listed in Table 2 above over time: from 1937 – present in the MMBR – and from 1923 – present in the JEB.

To establish a diachronic reference point for this evolution, we studied the four types of APCPs in a mini-corpus composed of journal articles and books on microbiology, cell biology, biochemistry, bacteriology, and genetics dating from the 1860s to the 1910s (list provided in Annex B). Of the four types of compounds studied, the most frequent were possessive compounds (*rod-shaped, short-chained, dark-colored, white-furred, single-layered, single-toed, thick-walled*), which nevertheless have few occurrences (approximately 400 in the whole corpus), followed by a significant number of -Ving PCPs (*chlorophyll-bearing, nitrogen-holding, iron-producing:* about 250 in all). The corpus contains very few occurrences of -Ved PCPs and APCs. Leaving aside lexicalised compounds in which the participle is modified by an adverbial (*well-known*, *well-developed*, *well-defined*, *deep-seated*), we found only 70 occurrences of *sex-limited* and 2 occurrences of *fluid-filled*. As for ACPs, leaving aside *-free* forms (15 occurrences), the only compounds we encountered were *water-soluble* (4 occurrences), *(toxin/arsenic/serum)-resistant* (16 occurrences) and *(mutually/immediately/directly)-dependent*. We checked whether a non-hyphenated alternative of APCPs exists in the corpus and found no form with the exception of *sex controlled characteristics.* These statistical data show that while the mechanism for forming APCPs was available, it was not "profitable" (Bauer 2001: 211), and not yet the distinctive characteristic of the scientific register which it is today.

3.3.1 Raw frequency

When raw (token) frequencies of APCPs are considered, the increasing importance of this form of syntactic re-packaging appears obvious for some of the bases considered. In the JEB corpus, compounds of the type *-free*, *-specific*, *-dependent*, *-sensitive*, *-rich* evolved from 72, 5, 0, 1 and 0 occurrences respectively in the 1920s to 137, 8, 1, 12 and 6 occurrences in the 1930s and to 1822, 8217, 5599, 3148 and 2177 occurrences in the 2000–2010 period. Where -Ved PCPs are concerned, the frequency of compounds such as *-mediated*, *-induced*, *-controlled* or *-treated* increased from 0, 4, 1, 3 occurrences in the 1920s to 2096, 5940, 613, 1471 occurrences in the 2000–2010 period.

Obviously, raw frequency figures are not significant by themselves. The size of the articles published by the two journals as well as the number of articles per issue and the number of issues per year also increased over time. As the total volume of text published by the two journals increased, it is natural that the number of compounds should have increased accordingly. However, the surge in ACP and PCPs cited above exceeds by far an increase proportional to the size of text published by decade. We estimate that the MMBR published 8–10 times more material in the 2000s than in the 1930s–1940s and that the JEB published around 20 times the volume of text it published in the 1920s. For most of the compounds cited, the increase in frequency is clearly far beyond the frequency projected from the increase in the published text volume. Given the increase in the volume of published text, compounds like *-specific* or *-induced* might have been expected to have about 160 and 80 occurrences respectively in the JEB throughout the 2000–2010 decade, while their actual frequency is 50 and 74 times higher respectively. This alone may serve as tangible evidence for the importance that the phenom-

enon has gained. In order to better estimate the extent to which compound pre-modification has gained ground, we computed three extra measures on a randomly selected sample of articles taken from the MMBR and JEB journals: (1) the normalised frequency of each particular type of compound (per million words), section 3.3.2, (2) the percentage of this particular type of compound among all hyphenated compound forms (section 3.3.3) and (3) chi-square values of the relative token frequencies taking into account the sample corpora sizes in the 1950s and the 1990s (section 3.3.4).

But before moving on to these measures, a few more observations may be made based on raw frequency figures. First, a diachronic study of token frequency of these compounds gives an indication of the approximate time when one particular type of compound, relatively frequent in the decade 2000–2010, was first used.

In some cases the late introduction of a particular form of compound may be correlated with the late introduction of (a particular use of) the adjective or the verb heading the compound. This is the case with the adjective *competent*. The specialised use of the adjective *competent*, meaning "having the capacity to respond (as by producing an antibody) to an antigenic determinant" (cf. Merriam-Webster Medical) was first used in the MMBR in the 1950s, i. e. one decade before the compound was first introduced. Similarly, ACPs such as *reactive to/with* (*virus-reactive, thiol-reactive,* etc.), *tolerant to* (*salt-tolerant*, *penicillin-tolerant*, *solvent-tolerant*) or PCPs: *linked to/with* (*enzyme-linked*, *telomere-linked*, *ester-linked*), or *driven by* (*light-driven*, *respiration-driven*, *sodium-driven*) were introduced one decade after their heads.

In other cases, the compounds seem to have been introduced much later than their heads. *-coupled* was introduced the 1960s, even though the head was used as early as the 1930s in the MMBR. Similarly, in the JEB, *-directed*, *-associated* or *-susceptible* were introduced much later (1960s–1970s) than their heads, which were already used in the 1920s. A great shift in the productivity (understood as profitability) of APCPs seems to have taken place in the 1960s–1970s, as shown by the evolution of the normalised frequency of most of the compounds studied (figures 2 to 5). We believe that during this period, using APCPs in scientific articles became a "fashion", which Plag (2003: 60) defines as the most important pragmatic, usage-based factor influencing profitability: "extra-linguistic developments in society […] make certain words or morphological elements desirable to use".

3.3.2 Normalised frequencies

Further evidence of the increase in the productivity of APCPs comes from another measure computed on samples of articles from the MMBR and the JEB collected for each decade: normalised frequency (PMW). Charts tracking the evolution of the PMW frequency of sample compound heads studied all show that most compounds of this type became more frequent over the decades, with the exception of *-free*, in both the MMBR and JEB, be they headed by adjectives (figures 2 and 3) or participles (figures 4 and 5). The exception represented by *-free* may be correlated with the increase in its occurrence in English for general purposes.

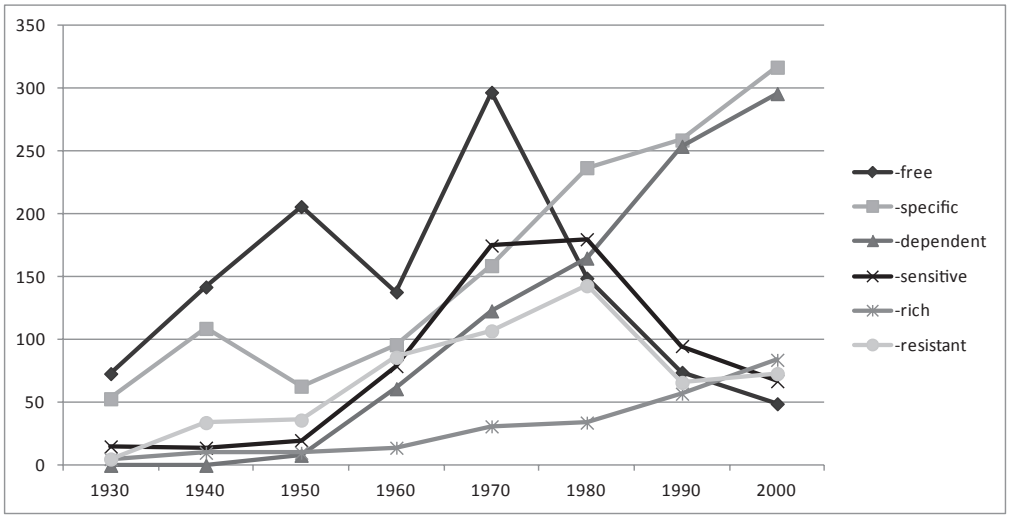

Fig. 2: Evolution of the PMW frequency of ACPs in MMBR

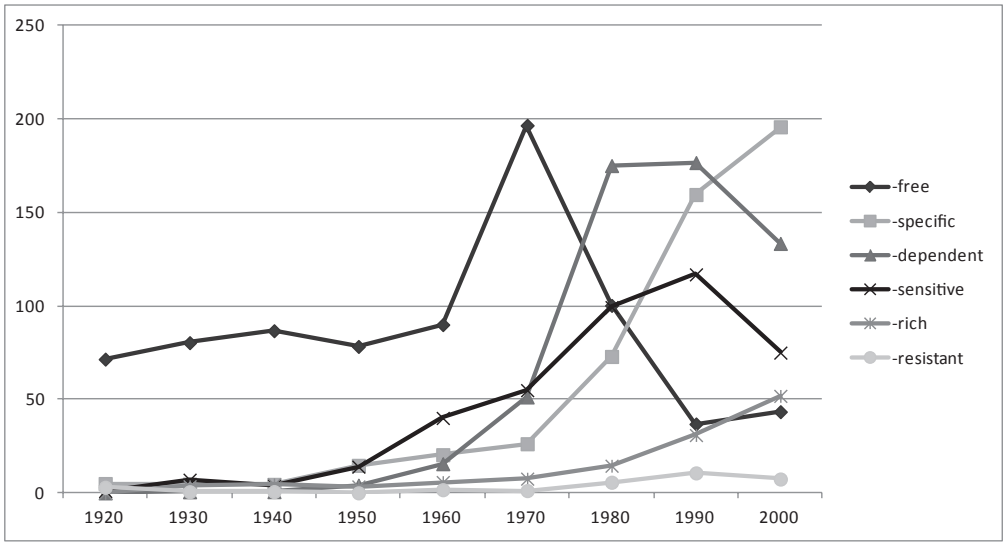

Fig. 3: Evolution of the PMW frequency of ACPs in JEB

The evolution curve for -Adj compounds is very similar in both journals. Apart from *-free* compounds, all have higher PMW frequencies in 2000 than in the 1920s. Some curves reach a peak around the 1970s–1980s, then slightly decrease, others grow continuously, while others remain relatively low, even though they slightly increase.

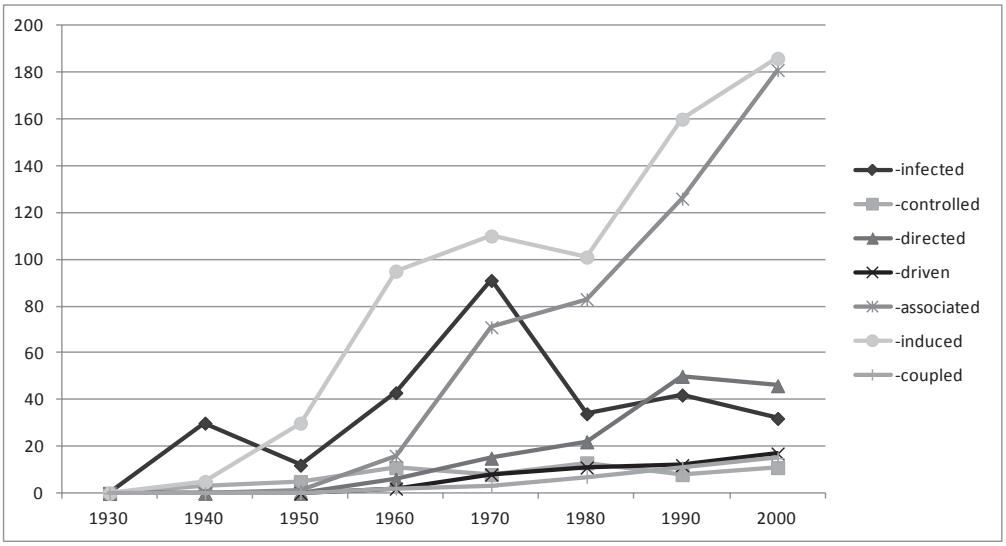

Fig. 4: Evolution of PCP PMW frequencies in the MMBR

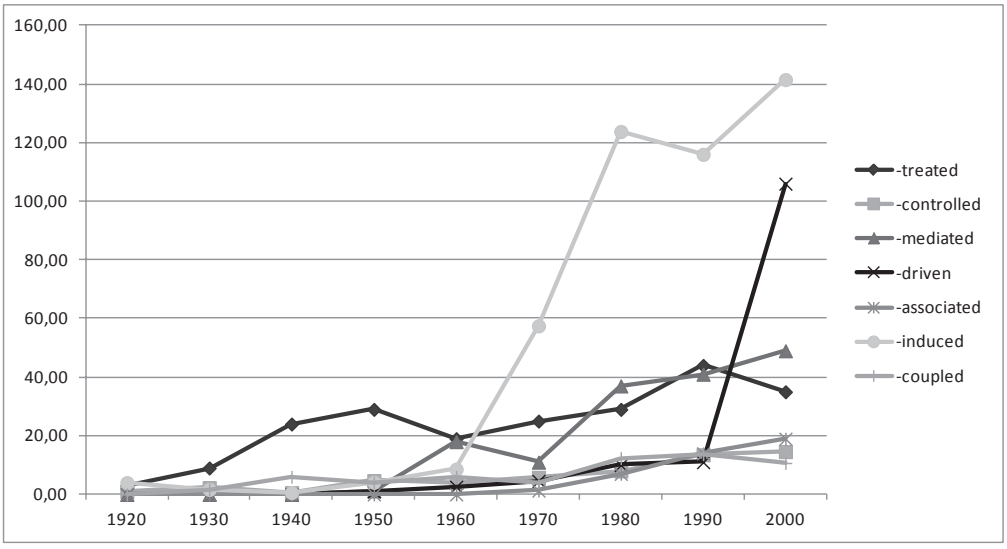

Fig. 5: Evolution of PCP PMW frequencies in the JEB

When the 22 types of APCPs studied are considered together, it appears clear that the moment when their productivity greatly increased can be situated somewhere in the 1960s–1970s (figure 6 illustrates this process in the MMBR).

Fig. 6: Evolution of the 22 compounds studied

3.3.3 Variations in percentage among all hyphenated compounds

The evolution tendency suggested by the data on PMW frequencies is confirmed when we consider the frequency of APCPs among all hyphenated compounds. We computed this percentage to isolate the evolution of APCPs from the evolution of hyphenated compounds in general. The frequency of the 22 types of APCP heads studied was found to have dramatically increased from the 1920s to the 2000s. Thus, the proportion of ACPs increased from 7.8 % to 17.3 % in the MMBR and from 2.6 % to 8.5 % in the JEB. Similarly, -Ved PCPs increased from 0.09 % to 15.04 % in the MMBR and from 0.7 % to 5.01 % in the JEB. While for certain forms we recorded a very slight increase or even a slight decrease (*-soluble*, *-free*, *-rich*, *-competent*, *-treated*, *-infected*), others increased by more than 1 %: *-specific*: +1 %, *-sensitive*: +1 %, *-resistant*: +1.3 %, *-rich*: +1.5 %, *-bound*: +1.6 %, *-mediated*: +3.2 %, *-associated*: +3.3 %, *-induced*: +3.4 % in MMBR.

3.3.4 Chi-square test for comparing proportions

In order to show that the increase in frequency of APCPs over time is not accidental or a result of the increase in the number of issues, number of reviews published per issue and the average size of reviews, we built two sample sub-corpora of JEB reviews in order to represent two decades remote from each other: the 1950s (2 million words) and the 1990s (15 million words). To be able to state that, taking into account the variations in the corpus sizes, the increase in the frequency of a given ACP or PCP base is significant, we ran the chi-squared test (the null hypothesis being that the frequency of an APCP is independent of the decades, and the predetermined level of significance being 0.05). The large chi-square values which we obtained, and extremely low p-values, far below the level of significance, indicate that the observed results

would be highly unlikely under the null hypothesis: the discrepancy between the results is unlikely to be the result of chance alone. Table 3 gives these values for several ACP and PCP bases.

Tab. 3a and 3b: Chi-square and P-values over proportion of APCPs in the 1950s and 1990s

ACP	1950s	1990s	Chi-sq	P-value
-soluble	84	78	336.59	< 2.2e-16
-free	229	584	136.36	< 2.2e-16
-specific	43	2529	385.64	< 2.2e-16
-dependent	12	2800	492.97	< 2.2e-16
-sensitive	41	1852	264.67	< 2.2e-16
-rich	10	490	71	< 2.2e-16
-resistant	0	169	31	2.34e-008
-reactive	3	28	0.9	0.34
-tolerant	1	222	38.99	4.26e-010
-dense	0	121	22.32	2.29
-competent	0	61	11.25	0

PCP	1950s	1990s	Chi-sq	P-value
-linked	4	156	21	3.84e-006
-mediated	3	650	114.02	< 2.2e-16
-associated	0	232	42	6.02e-011
-driven	3	183	28	1.19e-007
-induced	12	2563	449.26	< 2.2e-16
-directed	1	128	21	3.25e-006
-controlled	14	215	16.6	4.57e-005
-coupled	12	218	19	8.50e-006
-treated	85	706	15	7.27e-005
-infected	1	16	1.2	0.25

4 Conclusions and further work

The present work has shown that APCPs are significantly more frequent in Scientific English than in EGP. This hypothesis was tested using reliable statistical data by comparing two reference corpora for English (BNC and CoCA) with two specialised corpora (the PLoS corpus and the STEP corpus). We have also shown that the APCP-building mechanism is fully productive in the specialised genres and fields of knowledge we studied. Finally, a diachronic study carried out on two scientific journals which appeared early in the 19th century: MMBR and JEB showed that the mechanism for producing APCPs has become gradually more productive over

the decades. While certain forms appeared for the first time during the 1930–1950s, they have become very frequent during the last 3–4 decades.

Further research may be devoted to studying the co-existence of APCPs with the syntactic structures from which they were derived. A collocational analysis could compare the frequency of the adjectives/participles in APCPs and in other contexts. This type of analysis could be carried out both synchronically, in corpora such as PLoS or STEP, and diachronically in the JEB corpus.

Further research may equally be devoted to the translation of these structures into languages which do not make regular use of APCPs, such as French and other Romance languages. While English syntactic rules allow the attributive use of phrasal elements in order to shorten sentences and condense the noun phrase, different strategies have to be adopted when translating to Romance languages, reflecting their different morpho-syntactic structure. Several studies (Chuquet/Paillard 2007 for V-ing PCPs, Maniez 2010 for ACPs in the medical field) have been devoted to the difficulties in the translation of APCPs into French. We believe that study of the translation of APCPs in the two fields of knowledge illustrated in this article would be an interesting contribution.

References

Adams, Valerie (1973): *An introduction to modern English word-formation*. (English language series 7). London: Longman.

Adams, Valerie (2001): *Complex Words in English*. Harlow: Pearson Education/Longman.

Baayen, R. Harald (1993): "On frequency, transparency, and productivity." *Yearbook of Morphology 1992*. Eds. Geert E. Booij/Jaap van Marle. Dordrecht: Kluwer Academic Publishers. 181–208.

Bauer, Laurie (1983): "Compound Adjectives." *English Word-formation. Cambridge Textbooks in Linguistics*. Cambridge: Cambridge University Press. 209–213.

Bauer, Laurie (2001): *Morphological Productivity*. Cambridge: Cambridge University Press.

Bauer, Laurie (2010): "The typology of exocentric compounding." *Cross-disciplinary issues in compounding*. Eds. Sergio Scalise/Irene Vogel. (Amsterdam studies in the theory and history of linguistic science. Series 4, Current Issues in Linguistic Theory). Amsterdam/Philadelphia: Benjamins. 167–176.

Bauer, Laurie/Renouf, Antoinette (2001): "A Corpus-Based Study of Compounding in English." *Journal of English Linguistics* 29.2: 101–123.

Biber, Douglas (1988): *Variations Across Speech and Writing*. Cambridge: Cambridge University Press.

Chuquet, Hélène/Paillard, Michel (2007): «Les adjectifs composés en X+V+-*ing*: prédication, collocations, traductions.» *Palimpsestes* 19: 13–34.

Conti, Sara (2006): *Compound Adjectives in English. A Descriptive Approach to Their Morphology and Functions*. PhD, University of Pisa, Italy. 01.09.2014 <https://etd.adm.unipi.it/theses/available/etd-07142007-125640/unrestricted/TESI_SARA.PDF>.

Cortés, Jésus Andrés/Rigol Verdejo, Rafael (2006): "Teaching How to Generate Complex Nominal Phrases Used to Define New Concepts or Describe Objects." *Actas de V Congreso Internacional AELFE [Archivo de ordenador] = Proceedings of the 5th International AELFE Conference*. Ed. Claus-Peter Neumann, Ramón Plo Alastrué, Carmen Pérez-Llantada Auría. CD-ROM.

Francis, Gill (1994): "Labeling Discourse: an Aspect of Nominal-Group Lexical Cohesion." *Advances in Written Text Analysis*. Ed. Malcolm Coulthard. London/New York: Routledge. 83–101.

Gotti, Maurizio (2008): *Investigating specialized discourse*. Bern et al.: Lang.

Halliday, Michael Alexander Kirkwood/Martin, James Robert (1993): *Writing science literacy and discursive power*. London: Falmer Press.

Hart, David (1994): "Compound Forms. Description and Classification, with Particular Reference to Compound Adjectives." Historical English Word-Formation: Papers Read at the Sixth National Conference of the History of English: University of Urbino, 24th and 25th September 1993. Ed. Rolando Bacchinelli. Urbino: QuattroVenti.

Hopper, Paul J./Traugott, Elizabeth (2003): *Grammaticalization*. Cambridge: Cambridge University Press.

Kastovsky, Dieter (1992): "Semantics and Vocabulary." *The Cambridge History of the English Language*. *Vol. I: The Beginnings to 1066*. Ed. Richard Hogg. Cambridge: Cambridge University Press. 290–408.

Kastovsky, Dieter (1995): "The syntactic aspects of word-formation – Where are we today?" *Studies in Anglistics.* Eds. Gunnel Melchers/Beatrice Warren. (Stockholm Studies in English LXXXV). 157–179.

Lieber, Rochelle (1983): "Argument Linking and Compounds in English." *Linguistic Inquiry* 14.2: 251–285.

Lieber, Rochelle (2004): *Morphology and Lexical Semantics.* Cambridge: Cambridge University Press.

Ljung, Magnus: (2000): "Text Condensation in the Press: The Case of Compound Adjectival Premodifiers." *Words: structure, meaning, function*. Eds. Christiane Dalton-Puffer/Nikolaus Ritt. Berlin/New York: Mouton de Gruyter. 205–216.

Maniez, François (2010): «La traduction des adjectifs composés en langue médicale : étude d'un corpus bilingue anglais-français». Seminar held at CRTT, University Lyon 2.

Marchand, Hans (1969): *The Categories and Types of Present-Day English Word-Formation*. München: C. H. Beck.

Meys, Wilhelm Johannes (1975): *Compound adjectives in English and the ideal speaker-listener*. Amsterdam: North-Holland Publishing.

Opizzi, Alessia (2006): *Compounding in specialized languages*. Trento: UNI service.

Ormrod, Janet (2004): "Creation and subsequent usage of terms for discourse purposes in scientific research articles." *Actes de GLAT, Barcelona*, Universitat Politecnica de Catalunya, Spain.

Plag, Ingo (2003): *Word-formation in English.* Cambridge: Cambridge University Press.

Plag, Ingo (2006): "Productivity." *The Handbook of English Linguistics.* Eds. Bas Aarts/April McMahon. Malden/Oxford: Blackwell. 537–557.

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

Roeper, Thomas/Siegel, Muffy E. A. (1978): "A Lexical Transformation for Verbal Compounds." *Linguistic Inquiry* 9.2: 199–260.

Selkirk, Elisabeth O. (1982): *The syntax of words.* (Linguistic Inquiry Monographs 7). Cambridge: MIT Press.

Wälchli, Bernhard (2005): *Co-Compounds and Natural Coordination*. Oxford: Oxford University Press.

Annex A: Specialised Dictionaries

Dictionary of DNA and Genome Technology, 2nd Edition (2010). Ed. Paul Singleton. Chichester: Wiley.

Oxford Dictionary of Biochemistry and Molecular Biology (2006). Oxford: Oxford University Press.

Dictionary of Microbiology and Molecular Biology, 3rd Edition (2006). Ed. Paul Singleton/Diana Sainsbury. Chichester: Wiley.

Dictionary of Cell and Molecular Biology, 4th Edition (2007). Ed. John M. Lackie. Amsterdam et al.: Elsevier.

Annex B: Specialised Monographs dating 1864–1914

Bateson, William (1913): *Problems of genetics.* New Haven: Yale University Press.

Calkins, Gary Nathan (1914): *Biology*. New York: H. Holt & Co.

Carey, George Washington (1894): *The Biochemic System of Medicine*. St Louis: F. August Luyties.

Deason, J. (1911): *Practical Bacteriology*. Kirksville: Journal Printinc Co.

Dendy, Arthur (1912): *Outlines of Evolutionary Biology*. London: Constable & Co.

Journal of Genetics. Vol. 1 (1910–1911), Vol. 2 (1912–1913), Vol. 3 (1913–1914). Cambridge: Cambridge University Press.

Lock, Robert Heath (1906): *Recent progress in the study of variation, heredity, and evolution.* London: J. Murray.

Moore, Veranus Alva (1912): *Principles of microbiology: a treatise on Bacteria, Fungi and Protozoa pathogenic for domesticated animals*. Ithaca: Carpenter & Co.

Spencer, Herbert (1864): *The Principles of Biology.* Vol. 1. London: Williams and Norgate.

Walter, Herbert Eugene (1914): *Genetics. An Introduction to the Study of Heredity*. New York: Macmillan.

Alexandra Mestivier-Volanschi
Centre de Linguistique Inter-langues, de Lexicologie,
de Linguistique Anglaise et de Corpus
UFR des Etudes Interculturelles et Langues Appliquées (EILA)
Université Paris Diderot
Bâtiment Olympe de Gouges, case 7002
8 place Paul Ricoeur,
F–75013 Paris, cedex 13
avolansk@eila.univ-paris-diderot.fr