



Nominal compounds and their collocates in scientific discourse: a corpus-based analysis and pedagogical implications

Khadoudja Belkhenchir

► To cite this version:

Khadoudja Belkhenchir. Nominal compounds and their collocates in scientific discourse: a corpus-based analysis and pedagogical implications. *Etudes en didactique des langues*, 2013, *Quelle grammaire en LANSAD ? / What grammar for ESOL?*, 20, pp.113-129. hal-04046982

HAL Id: hal-04046982

<https://ut3-toulouseinp.hal.science/hal-04046982>

Submitted on 27 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Nominal compounds and their collocates in scientific discourse: a corpus-based analysis and pedagogical implications

Khadoudja BELKHENCHIR

Maitresse de Conférences

Laboratoire LOAPL – Université d'Oran



Whereas in literature and especially in poetry, new words are complex, in science they are compound (Salager, 1984: 141).

Introduction

The present article investigates how language functions in extremely specialised texts, focusing mainly on grammar. It specifically looks into how nominal compounds (henceforth NCs) are expressed in English and how they affect text interpretation and comprehension.

NCs are a general characteristic of English word formation in both general and scientific discourse. They may be simple (two nouns) or complex (three nouns and more). This latter type causes a real comprehension problem to non-native students (in our case, Algerian undergraduate learners) as these students are sometimes unable to find out what refers to what, for “word combination conveys an important load of information concentrated in one nominal group as this determines efficiency in the subject field” (Thouvenin, 1996: 33). The subject-field under study in this article is biology. In biology discourse, scientific concepts are often expressed through multi-word segments. To verify such a statement, a corpus-based analysis of authentic texts, *i.e.* fifteen biology articles from different scientific journals (British and American) was carried out using a computational programme called ATA (Aston Text Analyser). We will focus on compound nouns used in biology texts by presenting their frequency of occurrence and their collocates, together with their forms, and the different functions that they perform. We will then suggest some activities and tasks to teach NCs to non-native speakers of English and demonstrate how to raise these learners’ language awareness. By combining the study of scientific texts with corpus data, we aim “to make the learners aware of recurrent patterns of use” (Chambers, 2007).

Reasons for dealing with NCs

The study of the fifteen biology journal articles showed the complexity of the use of NCs due to:

- their frequency of occurrence and their length because it was noticed that “the more specialized the text, the longer the CNPs [Compound Noun Phrases]” (Salager, 1984: 142);
- their rarity in other languages such as French and Spanish (two Romance languages) and Arabic (a Semitic language known for its nonconcatenative morphology). In Spanish, for example, Carrio Pastor states that,

CNPs present difficulties for Spanish learners of English that are the result of the lack of rules to guide them and the different procedures used in a foreign language to express the same reality. English can simply juxtapose different words by their semantic relationship, as in “blood urea nitrogen concentrations.” However Spanish is not a synthetic language and needs connectors to join the different elements of a phrase, *i.e.* “concentraciones de nitrogeno en la urea de la sangre” (Carrio Pastor, 2008: 30).

For Algerian undergraduate learners of English, scientific discourse complexity presents a problem for two main reasons.

a. The nature of the language used as a medium of instruction

At school, Algerian learners are educated through Arabic and/or French, two languages which “would resort to a number of different propositions and relative clauses” (Salager, 1984: 140) instead of using long nominal compounds.

b. The lack of students’ exposure to English in general, and to scientific English in particular.

Besides foreign TV programmes, the only place where our learners hear, read or speak English is in the classroom (for one and a half to two hours per week, during six or seven years of their educational life). As undergraduate students, they have grasped only some basic English to survive in their educational field. They have some familiarity with some lexis and syntax of biology discourse, but they have never been exposed to such sophisticated English language as that of journal articles which they have to read to write up their project. So, when they read journal articles full of NCs, their problem is obviously greater than for native speakers.

But, if these structures which cause problems to non-native speakers, and sometimes even to native speakers of English, are still frequently used in scientific and technical English, there must be reasons, which are explained below.

Rationale behind the use of NCs

Through text study of the biology corpus, it was found out that NCs are used for three main reasons.

a. Economy

Both scientists and linguists agree that NCs are used for reasons of economy. This point is highlighted by Trimble (1979: 2) who finds that “English scientific and technical compounds usually represent ‘shorthand’ versions”. Halliday & Martin (1993:168) state that “the point of both the technical terms and the grammar is to compress as much information as possible into a short space” and that “without this condensation scientific texts would become very long, and probably unreadable, even for professionals” (*ibid.*: 172). In the same line of thought, Salager (1984: 140) notes that “numerous compounds prove more economical than the corresponding phrases which would use many function words.”

b. Scientific meaning

Nominal compounds (although generally composed of everyday words) form a lexis that is specific to the discipline in which it is used. Halliday & Martin remark that “technical language both compacts and changes the nature of everyday words” (1993: 172), that is general language words that become “terminologised” (Sager *et al.*, 1980: 242), and thus generate a new meaning. A comparison made by Knowles is quite relevant to this point: “Words are like chameleons. Change their environment and they change their colour” (Knowles, 1995: 10). Therefore, some everyday words *e.g.* base and pair, used in a specific subject-speciality (*e.g.* biology) become the scientific term “base pair”.

Accordingly, Salager states that “CNP corresponds to new terms, to some kind of technical jargon” (1984: 140) and finds that “the CNP (Compound Noun Phrase) is crystallized into a fixed expression owing to scientific meaning which the individual constituents do not have” (1984: 142), and “the difficulty usually lies in decoding the compounds rather than in understanding the individual words in the compound” (Master, 2003: 2).

These individual constituents “chunk” or “collocate” together, hence generating a new meaning. This process forms the third reason for the use of NCs in scientific texts.

c. Collocation

“Collocation”, where NCs tend to “chunk” and show strong bonding between the different constituents, is used in most scientific discourse, and biology is no exception. However, these NCs form a jargon which our undergraduate biology students have not grasped yet, and

[...]when three or more nouns are in immediate sequence, we encounter *the problem of the direction of modification*, in other words, what modifies what, or what is referred to what (Bartolic, 1978: 270).

In addition to the difficulty mentioned above (length of NCs), the frequency of occurrence of NCs in biology texts is also seen to present an obstacle for non-native speakers of English. The more frequently NCs are encountered in texts, the more our learners have to struggle to understand the text.

Materials and method

1. The corpus

The examples of English NCs in biology discourse used in this article are taken from a corpus consisting of a collection of authentic texts representative of the language of biology *i.e.* fifteen journal articles dealing with biochemistry and genetics, and counting seventy-six thousand running words/tokens. These articles were published between 1985 and 1994¹ in scientific journals such as *Biochemistry*, *Nature*, *Peptides*, *Biogenic Amines*, *Neuropharmacology* and *Naunyn-Schmedeberg's Archives of Pharmacology*.

All the articles were read by Algerian final year biology students during the academic year 1994-1995 at the writing-up stage of their end-of-course project *mémoire*.

2. The analysis of the biology discourse

The biology discourse study shows the grammatical and lexical features used in this corpus:

- grammar: a frequency count of tenses, voice (active/passive), adjectives, participles and modals;
- lexis.

In the present article, we are concerned with the biology lexis in the form of NCs.

3. The lexis of the biology discourse

The lexis used throughout the biology texts is made up of:

¹ The corpus, although outdated, is used in this article because:

- we aim to refer to authentic materials (scientific journal articles *read by students* (not chosen at random);
- scientific language has not changed since then (see footnote 2);
- I do not teach biology students any more to be able to select another authentic corpus.

a. **general language words**, some of which are common to all the authors, especially when it comes to presenting results or describing an experiment, for example: “As table 5 *demonstrates*”; “our results *suggested* (experiment 1) and subsequently *confirmed* (experiment 2) that...”, “This study *tested*...”, “similar results *were reported*”, etc.;

b. **specialised language**

Specialised biology language is of three types:

- *scientific terms* consisting of a single nominal and these are mainly derived from Latin and Greek origins and are almost universally used *e.g.* in English (bacteria, hormone, etc.), in French (bactérie, hormone) and in Arabic (بكتيريا) and (هرمون). This means the same word (sometimes with differences in the phonological system) is used in different languages. These words seem to present no major difficulty for our learners as they are common to almost all languages;
- *acronyms* like PCR (Polymerase Chain Reaction), APP (Amyloid Protein Precursor) and abbreviations (“DA” for “dopamine”);
- *scientific terms* consisting of a nominal group compound (plasma ketone bodies, soybean amino acid mixture).

4. NCs in biology discourse and their frequency

Biology discourse used in journal articles is significantly characterised by the widespread use of NCs. As Halliday and Martin (1993: 172) put it “science could not be science without deploying technical discourse as a fundamental tool” and Phillips (1985: 31) stated that “the choice of vocabulary in a text is largely a function of the subject matter”.

In a biology context, the “choice of vocabulary” is based on highly scientific terms presented in noun strings (NCs). Therefore, because scientific journal articles published in English address experts (and biology journal articles are no exception), they use a language for the expert which causes severe processing problems for lower-proficiency learners *i.e.* for our learners who are not experts in biology yet, and who are non-native speakers of English. The linguistic form of NCs is rarely used in their L1 (Arabic and/or Berber) and if used, it reads differently from English, and is absent in their L2 (French). For example, in French, we say “concentration de cholestérol dans le sang” for “blood cholesterol concentrations”, and “échantillons de sang” for “blood samples”, where “*de*” meaning “of” is always used. In Arabic, we also use prepositions, and we say: ارتفاع الكولسترول في الدم for “Blood cholesterol concentrations”.

Before dealing with the whole biology corpus, one text has been selected to illustrate the widespread use of NCs in biology discourse. It is entitled “Hypercholesterolemic responses in rabbits to selected groups of dietary

essential amino acids” (1994: 364-370) by E.M. Kurowska & K.K. Carroll from the American Institute of Nutrition.

The length of the NCs used in this text varies between two-noun NCs and five-noun NCs.

Number of constituents of NCs	Number of occurrences of NCs in the text	Examples
2 nouns	34	<i>base pair – plasma cholesterol – blood samples – animal proteins – plant proteins – ketone bodies</i>
3 nouns	12	<i>plasma ketone bodies – blood cholesterol concentrations – plasma cholesterol concentrations</i>
4 nouns	5	<i>casein amino acid mixture – casein amino acid diet</i>
5 nouns	1	<i>soybean protein amino acids</i> (where <i>soybean</i> is itself a NC formed of 2 nouns, <i>soya</i> and <i>bean</i>)

Table 1 - Number of constituents and occurrences of NCs in one text

The examples given in *Table 1* are not only NCs as such, but they form collocations where, for example, “base” always collocates with “pair”, “cholesterol” collocates with “plasma” or “blood”, “casein” with “amino acid”, etc.

Analysis and results

The biology corpus study deals with two aspects of NCs: forms and functions.

- *The forms* of NCs include NC specification and the number of its constituents *i.e.* NC length.
- *The study of functions* of NCs in scientific discourse is based on Bartolic’s categorisation (1978).

1. Forms of NCs

Besides their frequency and length, NCs show a considerable variety of forms. The latter can also add another feature of complexity for our learners as Williams (1984: 150) observes: “length [of NCs] correlates with formal and functional complexity”.

a. Nominal compound specification

NCs are formed of a headnoun (it is the final noun) preceded by a number of other noun modifiers. These, according to Bartolic (1978: 259) “can be freely

placed before a headnoun and their number is mainly restricted by the information that is to be conveyed”.

Nominal compounds, referred to in this article, are those elements (content words) made up of two individual constituents and more, separated one from the other by a space (“cell line”, “Ltk mouse fibroblast”, etc.) or a hyphen. The words separated by a hyphen are considered different elements in the NCs as in “dose-response curves”, “antigen-overlay method”. In these examples, the NCs were classified as being made up of three different elements.

Nominal compounds are those which start and finish with a nominal (“*blood cholesterol concentrations*”) and may sometimes incorporate an adjective (“*strand-specific primers*”) or a participle (“*mock-transfected cells*”).

b. NC length: procedure

Each time a nominal compound was encountered in a text, it was highlighted and then the number of its lexical constituents was counted. These ranged from two-noun NCs to six-noun NCs. 7-nouns NCs were very rare (4 examples in the whole corpus). Here are some examples of each kind of NC found in the 15 biology journal articles.

Number of nouns in NCs	Examples
2 nouns	<i>rotavirus antigens – cholera toxin – adenylate cyclase</i>
3 nouns	<i>blood cholesterol concentrations – density gradient ultracentrifugation</i>
4 nouns	<i>antigen-overlay immunoscreening method – rat liver cDNA library – tissue kallikrein-binding protein</i>
5 nouns	<i>rat liver expression cDNA library – rat tissue kallikrein gene family</i>
6 nouns	<i>on-line narrow-bore PTH-amino acid analyzer – APP 770 expression constructs</i> (where APP stands for 3 nouns: Amyloid Protein Precursor)
7 nouns	<i>rat liver cDNA expression library, full-length APP 695 expression construct, full-length APP 751 cDNA sequence.</i>

Table 2 – Examples of multi-noun NCs

c. NC forms

Besides their frequency and length, NCs show a considerable variety of forms. The most frequent ones are: two NCs (N+N: cholera toxin); three NCs (N+N+N: blood cholesterol concentration) etc., but biology discourse also uses other forms: noun + participle + noun / noun + adjective + noun, etc. and as such it can also add another feature of complexity for our learners.

Below are some authentic examples of the various forms of NCs used in biology discourse.

Forms of NCs	Examples
Present participle + Noun	<i>yawning behaviour – seeding density</i>
Noun + Present participle	<i>protein staining – acid sequencing</i>
Noun + Present participle + Noun	<i>urine-collecting bags</i>
Noun + Past participle + Noun	<i>filter-grown cells – intracellulose-coated sample foil – cholera toxin-induced diarrhoea</i>
Noun(s) + Adjective + Noun(s)	<i>cholesterol-free diets – adenylate cyclase dependent receptor sites</i>
Acronym + Noun	<i>CSF-brain barrier – APP secretase activity</i> (where APP stands for Amyloid Protein Precursor)
Abbreviation + Noun	<i>DA binding sites</i> (DA standing for Dopamine)

Table 3 – Forms of NCs

The preceding part shows the forms of NCs used in the biology texts. This was important to find out in order to suggest pedagogical solutions to overcome the difficulty caused by those elements involved in the formation of NCs.

Furthermore, NCs not only have a variety of forms, but their use is not arbitrary. They also perform different functions.

2. Functional analysis of nominal compounds

To speak about the different functions of NCs in scientific discourse, it is important to refer to the way scientists write their articles to understand why NCs have functions.

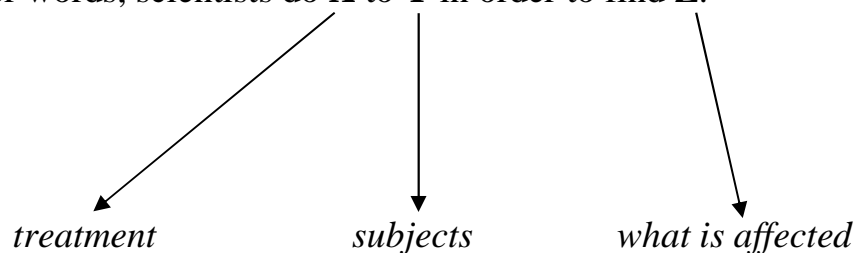
According to Myers (1990: 142), professional articles (the focus of this article) create what he calls “a narrative of science” *i.e.* they follow “the argument of the scientist, arrange time into a parallel series of simultaneous events all supporting their claim, and emphasize in their syntax and vocabulary the conceptual structure of the discipline” .

As stated differently by an Algerian biologist, before writing scientific articles, biologists start by doing the experiment and find results. Then, when writing the article, they refer to:

- a. the aim/objective of the experiment;
- b. the subjects used: animals (rabbits, mice, rats, etc.), plants (fungi, geranium, etc.) or human beings;
- c. the method used in the experiment:
 - diet,
 - treatment (specifying the instruments used),
 - kinds of instruments (syringes, scissors, dissection equipment, etc.),
 - sterilization of instruments to reduce the error rate of the experiment.
- d. external factors (temperature, light, ventilation, etc.);

e. They then prove or disprove the hypothesis.

In other words, scientists do **X** to **Y** in order to find **Z**.



This explanation of how biologists write their articles will lead us to the functional analysis of NCs. The latter has been dealt with by a number of authors such as Sager *et al.* (1980), Williams (1984) who actually used Bartolic's categorisation (1978) and Master (2003).

Bartolic's categorization, although related to a technical field, is suitable for our analysis of scientific discourse as the same functional categories apply to biology texts.

Bartolic's categorisation (1978) includes:

- operation including the following parts: operating principle, working substance in operation, means of operation, characteristic working part; person who formulated the operating principle;
- materials;
- purpose of application;
- location;
- professional engagement;
- shape/form.

These categories (Bartolic's), specific to a technical field, apply exactly to the field of biology.

Table 4 illustrates Bartolic's categories describing the functions of NCs completed by authentic examples selected from the biology discourse corpus. To Bartolic's categories, two more categories specific to the biology corpus have been added.

- Abbreviations or acronyms + headnoun' *e.g.*
Abbreviation + Noun: DA receptors (where "DA" stands for "Dopamine");
Acronym + Noun: PCR-based cloning method (where "PCR" stands for "Polymerase Chain Reaction") – APP secretase activity (APP: Amyloid Protein Precursor);
- The category "B of A" *e.g.* rat brain (brain of a rat), drug intake, stool loss, nucleotide exchange, etc.

Operation	
Operating principle	<i>amino acid analyzer – immobilon-N filter – elution buffer – rapid-quench apparatus</i>
Working substance in operation	<i>rice flour – glucose-electrolyte – milk formula – mycoplasma contamination – drug regime – snake venom – blood samples – amino acids – agarose gels</i>
Means of operation	<i>ear vein, etc.</i>
Characteristic working part	<i>weighing scale – syringe plungers – rapid-quench apparatus –</i>
Person who formulated the operating principle	<i>Wilcoxon rank – Ussing chambers – Ringer solution – Denhardt’s solution – Klenow fragment – Sprague-Dawley rats – Alzheimer’s disease – Parkinson’s disease</i>
Materials	
<i>wire bottoms – wire netting (about cages where mice under treatment are kept) – metal bars – filter paper</i>	
Purpose of application	
<i>treatment group – control group – dehydration solution – filter paper – control mice</i>	
Location	
<i>binding sites – cleavage sites – cell-lines – n-terminal region – coding region – enzyme subsites – camp-binding sites</i>	
Professional engagement	
<i>manufacturer’s instructions (for instruments, etc.)</i>	
Shape/Form	
<i>inhibition curves – loop structure – bell-shaped curve</i>	

Table 4 - Functional analysis of NCs following Bartolic’s categorization

The aim of the study of NCs in biology texts, besides their frequency, was to find out the forms and functions they present so that as teachers and researchers we can help “pedagogically” our learners to overcome the difficulty of recovering NCs so that comprehension of subject-speciality text follows.

The data found and the reasons explained above urge us to pay special attention to the teaching of NCs to Algerian learners of English in science departments. It is, therefore, important to deal with NCs in the biology course by developing different tasks to help learners understand how the English language functions in context. Unlike the native English reader, the non-native reader “has not internalised the set of recovery procedures that would help him, once transferred, operate on English NCs” (Williams, 1984: 146).

Pedagogical implications

As NCs frequency, their rarity in the readers’ L1 and L2 and their complexity, represent a comprehension problem for our learners, our immediate goal as EFL reading teachers is to minimise reading difficulties and to maximise text comprehension by presenting and explaining language patterning, focusing on one specific discipline, biology.

1. Computational analysis

Because lexis in biology discourse is formed of everyday words in the form of NCs that tend to “chunk” and “collocate” to give new meaning, a computational analysis is seen necessary as it will unveil all the language elements in the near environment of the NCs under study.

A specific computer programme called ATA (Aston Text Analyser)² is used for analysis. This latter is based on “concordancing” to find out how concepts are realised in scientific discourse and which grammatical structures these concepts combine with. As Stubbs (1996: 97) put it: “Meanings are conveyed directly, by the choice of particular words, but they are also conveyed indirectly by patterns of co-occurrence: which words collocate, and which words occur in which grammatical constructions”.

2. Steps for the preparation of concordance materials

The preparation of materials for concordancing follows necessary steps after having scanned and edited the texts in a computer-readable form.

a. Selecting items to concordance

It is the English teacher’s duty (after students’ consultation) to find out what items need to be selected and focused on.

The selection varies from text to text, and from what the students actually need to know *i.e.* what items, which principle, etc. need to be made clear and explicit.

The concordance lines do not themselves solve the problem of difficulty, but it is through the tasks that the students are asked to do that we hope to bring explanation and make the difficult items clearer.

b. Producing the concordance

The items to be concordanced are selected by the teacher focusing on whatever item s/he believes is relevant to whatever point s/he is teaching *i.e.* NCs.

Once the items to be concordanced have been selected, the teacher uses a concordancer to extract all the occurrences of that item or pattern and develop tasks accordingly, so that students can explore meaning and interpretation.

² ATA (although developed in the 1990’s) is used, instead of another more up-to-date concordancing programme, because the study results would have been the same as the observations on the scientific language have not much changed since *e.g.* a protein with its characteristics is still the same protein. Indeed, scientific jargons do not change quickly, and their validity often spans over generations. These mutations do not seem to have happened.

The example below gives a few concordance lines of the node “*blood*” from the biology corpus.

study, the protocol for	blood	sampling was similar, but
bodies were analyzed in	blood	samples collected in heparinized
the regulation of local	blood	flow and electrolyte transport
have variable effects on	blood	cholesterol concentrations. In ...

Figure 1 – Example of some concordance lines of the node “blood”

The concordance lines of the selected items are of the “Key Word In Context” (KWIC) type. It consists of the key word in the centre of each line (represented by an asterisk) with context on each side of it. With the ATA programme, there are four words preceding and four words following the KWIC, as shown above.

Also, if concordance lines are not clear enough to help students explore the meaning of the item under study, students can look at the surrounding text – the larger context (that is the sentence or the paragraph where the item is used) on the computer by clicking on “View context” from the ATA menu.

c. Editing the output

After the selection and extraction of different occurrences of the item are done, we should edit the output by eliminating unnecessary items (*e.g.* “hapax legomenon” *i.e.* one occurrence in a given corpus).

The steps described above are necessary for language lesson production as they form the basis for the development of different tasks.

3. The procedure

The teacher distributes an authentic biology journal article of a medium length (three or four pages) to read. It was thought that the only way to help our learners recover the meaning of NCs is to provide them with authentic materials where NCs are used in context. After reading the whole article, the following tasks are proposed:

a. Nominal compounds tasks

The teacher divides the article into different paragraphs, gives each pair/group of students a paragraph and asks them to interpret or decode NCs by various methods.

- *Locating* each of the NCs in the paragraph and turning the compound into a paraphrase.
- *Recovering* the meaning of each NC circled (by the teacher) in the paragraph by writing its expansion in the margin (using word order as a means of disambiguating *i.e.* to use the principle of right-to-left tracking). “Many students

whose native languages allow nouns to be postmodified by adjectives (*e.g.*, Spanish and French) need to be reminded that it is the final word in the English noun compound [...] that is the head noun” (Master, 2003: 5).

- *Asking* students to form NCs: the teacher gives the students phrases and asks them to turn the phrases into compounds *e.g.*

<p>Instructions: Change the words in italics into NCs.</p> <p><i>The concentrations of cholesterol in blood</i></p> <p><i>A diet based on amino acid of casein</i></p> <p>Researchers have located <i>the site for the binding of RNA</i></p>
<p>Answers</p> <p>Blood cholesterol concentrations</p> <p>Casein amino acid diet</p> <p>the RNA binding site</p>

- *Classifying* NCs according to their functions.

Instructions: Classify the following NCs according to the list of functions below.	
<ol style="list-style-type: none"> 1. Operation 2. Materials 3. Purpose of application 4. Location 5. Professional engagement 6. Shape/Form 	<ol style="list-style-type: none"> a. cholesterol-free diets b. cell-lines c. inhibition curves d. wire netting e. antigen-overlay immunoscreening method f. Denhardt’s solution g. loop structure h. binding sites i. urine-collecting bags j. Wilcoxon rank

- *Translating.*

Give the students noun groups in Arabic and/or French and ask them to translate them into NCs in English. The aim of such an exercise is to check whether the students have internalised the process by which we form and read NCs in English.

- *Concordancing.*

On the computer, the teacher suggests concordance lines and asks students to find out the meaning of the NC by looking at the surrounding text (the co-text).

Another possibility would be that the teacher can simply bring concordance printouts (in the form of handouts) ready to be used in the classroom. The students will first analyse lexical and/or grammatical patterns in examples of the items in a set of concordance lines, and they then complete tasks related to these items.

Some other sample tasks using concordances in teaching English could be of the following types:

- sentences with gaps;
- putting in order;
- finding the rule, *i.e.* deductive teaching;
- exploring differences between forms and/or functions of words;
- looking at synonyms and checking how much overlap there is;
- examining texts for cohesion, etc.

The teacher distributes a copy of the concordance of the item under study and explains to students how to interpret it.

The teacher assigns each student one or more of the items to prepare. Students (depending on class size) can have one item each to prepare or they may share (if there are more students than items).

As an example, the teacher distributes concordance lines of the node “blood” and asks the students to list all the items that collocate.

SYNOPTIC PROFILE									
blood									
3	fall	6	of	*	14	pressure	2	and	
3	of	4	in	*	2	samples	2	control	
3	presence	2	improved	*	1	Haloperidol	2	or	
		2	mg/day	*	1	cholesterol	2	was	

CONCORDANCE									
blood									
Cholesterol analysis					Blood samples were taken from				
**** Continuous					Blood pressure control has been				
study, the protocol for					blood sampling was similar, but				
the latter drugs, however,					blood pressure did not change				
pressure, CI 201-678 improved					blood pressure significantly in an				
of dizziness and improved					blood pressure; measurement of blood				
be also involved in					blood pressure regulation (Margolius, 19				
while a fall in					blood pressure was frequently noted				
and cause fall in					blood pressure.				
bodies were analyzed in					blood samples collected in heparinized				
the regulation of local					blood flow and electrolyte transport (
at doses <2 mg/day					blood pressure increases slightly. This				
0.5 to 1 mg/day. Blood					pressure has been measured				
lity,6 mucus secretion,7 mucosal					blood flow,8 epithelial restitution and				
that show fall of					blood pressure, CI 201-678 improved				
and dysentery (presence of					blood in stools); and treated				
the unwanted drop of					blood pressure at low dose.				
blood pressure; measurement of					blood pressure in a supine				
treatment; the presence of					blood, mucus, or pus in				
visit; consistency; presence of					blood, pus, or mucus), clinical				
have variable effects on					blood cholesterol concentrations. In sev				
the regulation of renal					blood now, water, and sodium				
patients with this symptom					blood pressure control was performed.				
the regulation of systemic					blood pressure and the development				
amount clearing towards the					blood. Haloperidol which easily crosses				

From the concordance lines, we notice that the node “blood” collocates with “samples”, “pressure” and “flow”, thus forming:

- two- noun NCs (“blood samples”, “blood pressure” and “blood flow”), and
- three- noun NCs (“blood pressure control”, “blood pressure regulation”).

The student's individual task is to decide what meaning the node "blood" has in their example(s), and to present the explanation orally to their pairs.

This task is considered as a grammatical and/or stylistic exercise as well as an interpretative one. As such we prepare the students with this specific lexis.

b. Tasks dealing with participles (or participial adjectives)

The teacher and students first discuss the formation of participles, and give their own examples of how participles work and what functions they perform. The teacher writes these examples on the board to elicit the formation and placement of the participles in a sentence. This acts as a structural review and prepares the students for the next task.

Another possibility could be a task example of "Find the rule": rather than tell students a particular rule, teachers can give students a set of concordances and ask them to discover what rules can be deduced from the evidence.

• Re-ordering the left or right context

To increase the students' understanding of the position of these participial adjectives in sentences, concordance output can be used as gap-fill and/or matching-up exercises.

Using a word processor, the teacher can delete for example, right or left collocates of a node, and ask the students to supply these.

The node "cells" has been selected from the corpus for such a task.

human embryonic kidney 293	cells	(1)..... with APP 695...
parameters of cl 19A	cells	(2)..... on HAHY Millipore ...
derivatives secreted by sf 9	cells	(3)..... with baculovirus expressing...
in (4).....	cells	the APP secretase is normally ...
Microsomes prepared from COS-1	cells	(5)..... with ...
... then examined the (6)	cells	to identify the cellular

Correct answers of the words deleted:

1- *transfected*; 2- *cultured*; 3- *infected*; 4- *cultured*; 5- *tranfected*; 6- *transfected*

• The teacher can also give the participial in its infinitive form and ask the students to give the *-ed* or *-ing* forms as in the node "protein".

kallikrein with the (<i>bind</i>)	protein	in both humans and
sequence of the (<i>purify</i>)	protein	and by sequence comparisons
segment of the (<i>translate</i>)	protein	sequence from the cDNA
as that of the	protein	(<i>purify</i>) from rat sera
for comparison, a	protein	(<i>immunopurify</i>) from rabbit kidney
a 26 428-Da acidic	protein	whose (<i>derive</i>) amino acid

• The exercise can also be lexical, *i.e.* the students can be asked to supply all the adjectives that might modify the node "protein", and then compare their information with all the adjectives that they can select from the concordance

lines. The most frequent adjectives in the concordance lines that are used to describe the node “protein” in the corpus are: mature, microsomal, translated, binding, derived, purified, encoded, etc.

- Another exercise can be grammatical. In the concordance lines of “protein” the teacher can highlight for example, the following words “showed/called/encoded” and ask the students to determine whether these are “verbs” or “participials”.

Furhermore, the rK10	protein	<i>showed</i> the same molecular
Part of a larger	protein	<i>called</i> the amyloid protein
are conserved in the	protein	<i>encoded</i> by rklk10. The ...

The above tasks are but some examples of what one may suggest for future applications.

Conclusion

Nominal compounds are frequently used in English scientific discourse for reasons of economy as they compress as much information as possible into a short space. They, thus, give domain specific vocabulary crystallized into a fixed expression which the individual constituents do not have. In contrast, NCs are not a common structure in many languages. For Algerian undergraduate students in science departments, who are taught through Arabic and/or French (two languages which would resort to the use of different propositions, relative clauses and prepositions), NCs’ complexity is a “real” problem.

In the present article, through a combination of scientific text study with corpus data, we have tried to describe NCs in their different states (kinds, length, forms) followed by their various functions (using Bartolic’s categorisation) to decipher the difficulties they may cause.

In terms of pedagogy, teachers should combine traditional language teaching with modern technological tools. Chambers (2007) recommends us to “study individual texts in detail as has traditionally been the case and complementing this with corpus data”. The tools should be easy enough to be used by both teachers and students.

Students should be provided with corpus data (concordance lines, etc.) prepared by teachers.

Teachers should be more flexible with non-native learners of English with NCs’ use. They should also provide and vary necessary language activities that allow learners to improve their domain specific linguistic competence.

We hope that this study will stimulate further investigation into other areas of NCs’ use and their interpretation for overall text comprehension.

Bibliographical references

- BARTOLIC, L. 1978. Nominal compounds in technical English. TODD-TRIMBLE, M., L. TRIMBLE & K. DROBNIC (eds.) *English for Specific Purposes: Science and Technology*, English Language Institute, Oregon State University, 258-277.
- CARRIO PASTOR, L. 2008. English complex noun phrase interpretation by Spanish learners. *RESLA (Revista española de lingüística aplicada)* 21, 27-44.
- CHAMBERS, A. 2007. Language learning as discourse analysis: implications for the LSP learning environment. *Asp* 51-52. <<http://asp.revues.org/483>;DOI:10.4000/asp.483>.
- HALLIDAY, M.A.K. & J.R. MARTIN. 1993. *Writing Science: Literacy and Discursive Power*. London & Washington D.C.: The Falmer Press.
- KNOWLES, F. E. 1995. Lexical cartography. *LSP texts*. Lexical studies file, Language Studies Unit, The University of Aston.
- MASTER, P. 2003. Noun compounds and compressed definitions. *English Teaching Forum*, July 2003. <<http://www.uhu.es/mpilar.ron/61109/units/unidad-2/resources/documents/noun-compounds.pdf>>.
- MYERS, G. 1990. *Writing Biology: Texts in the Construction of Social Knowledge*. Wisconsin: University of Wisconsin Press.
- PHILLIPS, M. 1985. *Aspects of Text Structure: An Investigation of the Lexical Organisation of Text*. Amsterdam: North-Holland Publishing Company.
- ROE, P. & F. E. KNOWLES. 1995. ATA programme. The University of Aston.
- SAGER, J.C., D. DUNGWORTH & P.F. McDONALD. 1980. *English Special Languages: Principles and Practice in Science and Technology*. Wiesbaden: Brandstetter.
- SALAGER, F. 1984. Compound nominal phrases in scientific-technical literature: proportion and rationale. PUGH, A.K. & J.M. ULIJN (eds.). *Reading for Professional Purposes: Studies and Practices in Native and Foreign Languages*. London: Heinemann, 136-145.
- STUBBS, M. 1996. *Text and Corpus Analysis, Computer-assisted Studies of Language and Culture*. Oxford: Blackwell.
- THOUVENIN, S. 1996. *The Identification and Exemplification of Multi-Word Units within a Technical Corpus of English, Including an Investigation of Nominal Groups*. MSc Dissertation, Language Studies Unit, The University of Aston.
- TRIMBLE, L. 1979. A rhetorical approach to reading scientific and technical English. *English Teaching Forum* 17 : 4, 2-6.
- WILLIAMS, R. 1984. A cognitive approach to English nominal compounds. PUGH, A.K. & J.M. ULIJN (eds.). *Reading for Professional Purposes: Studies and Practices in Native and Foreign Languages*. London: Heinemann, 146-153.