On alternating pre-modified and post-modified nominals such as *aspirin synthesis* vs. *synthesis of aspirin*: Rhetorical and cognitive packing in English science writing

Christopher Gledhill & Mojca Pecman

Abstract In this paper, we examine the alternation between pre-modified and post-modified nominals, such as *aspirin synthesis* as opposed to *synthesis of aspirin*. This type of alternation has been previously explained in terms of rhetorical function (whether the item is 'given' as opposed to 'new'), and phraseology (whether the item is seen as a 'packed' lexicalised term, or an 'unpacked' expanded nominal). In this paper, we suggest that other factors may also be involved, and we examine four specific cases from two different varieties of scientific English (Scientific Research Articles vs. Science Forums). Here we investigate the underlying constraints that govern this alternation, and we try to establish whether there is a preference for writers to 'unpack' nominal groups early on in a text and then to 're-pack' them later on. Overall, we suggest a number of parameters that may explain the choice of one structure over another. Finally, and more generally, we show that this grammatical variation is a particularly productive resource in English nominal groups, and as such contributes to the rhetorical and cognitive construction of scientific demonstration in particular and scientific discourse more generally.

Keywords binominals, distributional analysis, ESP, lexico-grammatical alternations, nominal compounds, multiple case study, scientific English

1 Introduction

This study focuses on a key linguistic resource in English science writing: the alternation between pre-modified and post-modified nominal groups. Our hope is that by examining this particular linguistic feature, we may more fully understand how meaning is constructed in English scientific and technical discourse. Over recent years, scientific and academic texts have become a key object of study for text linguists and discourse analysts; these researchers are interested in observing how new knowledge emerges through the analysis and interpretation of scientific discourse (Halliday 1998a, 1998b). As scientific and technical texts are produced in a highly-constrained discourse context and for a very specific community of experts, linguists who are interested in English for Specific Purposes (ESP) and Languages for Specific Purposes (LSP) have often focused on a few conventional features, such as the passive voice, modal verb usage, graeco-latin morphology, nominalisation of verbs, etc. However, we would suggest that more dynamic phenomena, such as those produced by alternations of nominal structures, also deserve the attention of descriptive linguists.

In this paper, we see the alternation between pre-modified and post-modified nominals not as an abstract grammatical transformation, but rather as an example of variation that takes place in running texts and dialogues, and which represents an essential resource for meaning-making in English. A particularly clear example of this can be seen in the following exchange between two science forum participants:[1]

[1] Source: The Science Forum 'The Naked Scientists' based at Cambridge University's Institute of Continuing Education, accessed 11 July 2015: <http://www.thenakedscientists.com/forum/index.php?topic=27009.0>.

(1a) What is the **aspirin synthesis** mechanism?
(1b) Essentially i'm looking for a proper curly arrow type mechanism for the **synthesis of aspirin** from salicylic acid (2-hydroxybenzoic acid) and ethanoic anhydride.
(1c) Curiosity strikes ... what *do* you want a mechanism for **aspirin synthesis** for, then??

This attested occurrence shows that there is a complex interplay between the first-mentioned nominal group (*aspirin synthesis*), an apparent alternation (*synthesis of aspirin*) and an apparent return to the original formulation (*aspirin synthesis*). In the following discussion, we call the pre-modified version of this nominal group (NG) 'packed' and we code this structure as <N1 N2> (i. e. *aspirin synthesis*), while we call the post-modified version 'unpacked' and we code its structure <N2 of N1> (i.e. *synthesis of aspirin*). Note that our use of numbers here does not imply that <N2 of N1> is 'derived' in any way from <N1 N2> or vice versa. Similarly, our use of the terms 'packed' and 'unpacked' does not imply that the composite <N1 N2> structure is necessarily anterior to <N2 of N1>, at least not in the cognitive or generative sense of 'base form' vs. 'derived form'.

In contrastive linguistics and translation studies, the choice between alternative forms of expression is of key significance. For instance, the alternation between <N1 N2> and <N2 of N1> occurs frequently in English, but is unknown in French. As teachers of specialised translation (especially French to English), we have observed that French students writing in English often hesitate between a pre- or post-modified nominal form. To briefly illustrate our point, here are a series of learners' translations from French to English, in which we encounter the alternating pair *aircraft withdrawal* vs. *withdrawal of aircraft*:[2]

Source text
1. Les programmes de **retrait des avions** bruyants
   Au plan international, depuis de nombreuses années, les avions les plus bruyants, caractérisés par des normes de certification énoncées dans le volume I de l'annexe 16 de la convention relative à l'aviation civile internationale, ont été progressivement retirés d'exploitation.

1.1 **Retrait des avions** classés en chapitre 2

Translation 1
1. Noisy **aircraft withdrawal** programs
   At the international level and for many years now, the noisiest aircraft, which are characterized by noise certification standards in Annex 16, Volume I to the Convention on International Civil Aviation, were gradually taken out of service.

1.1 Chapter 2 **aircraft withdrawal**

Translation 2
1. Noisy **aircraft withdrawal** programs
   At an international level, the noisiest aircraft, characterized by certification standards established in the Annex 16, Volume I of the Convention on International Civil Aviation, have been for many years progressively taken out.

[2] Learners' translations collected in the Master's course on specialised translation at the Department for Applied Languages of Paris Diderot University: http://www.eila.univ-paris-diderot.fr/enseignement/lea/master (for further details, see section 3 devoted to context and methodological aspects of investigating structural variation).

1.1 **Withdrawal of aircraft** classified in Chapter 2

Translation 3

1. Noisy **aircraft withdrawal** programs
   At an international level, since many years now, the noisiest aircraft, which are characterized by certification norms, formulated in the volume I of appendix 16 of the convention relating to the international civil aviation, have been withdrawn from operation.

1.1 **Withdrawal of aircraft** classified in chapter 2.

The problem here is that the students are expected to translate the same segment (*Retrait des avions*) in two different ways because, based on our experience and intuition, there appears to be an underlying tendency that makes the switch from <N1 N2> in the title to <N2 of N1> in the subsequent part of the text sound more natural. As far as we know, there is no clear evidence or rule in textbooks or research about this specific type of alternation. In Translation 1, the student has 'correctly' chosen the packed form for the first title, but he or she has 'incorrectly' chosen the same packed structure for the second title. Conversely, translations 2 and 3 have successfully chosen a packed form for the first title and the unpacked one for the second title. How can we explain our preference for Translations 2 and 3? The type of 'error' in Translation 1, if it is an error, is not easily perceptible nor classifiable according to the usual error typology. There is no deep syntactic error here, nor indeed an error related to terminology: instead, we – as the evaluators of these translations – just have a preference for one form in the initial part of the text, and later on in the same text a preference for the alternative form. Examples such as these show us that looking at students' translations can help to highlight the difficulties encountered when handling the very sophisticated linguistic process of alternation.

In the following sections, we explore the various linguistic and contextual factors that are involved in making this kind of subtle distinction between seemingly equivalent forms. In particular, we aim to show that the relative well-formedness and interpretability of this kind of alternation is dependent not only on rhetorical function (factors such as 'given' vs. 'new' information, degree of contrast, switch of focus and so on), but also phraseology (whether the sequence is a conventionalised or lexicalised sequence, or part of an expanded nominal group with post-modifiers and other elements). However, we also suggest that other factors may be involved, especially when one examines how a sample of terminological variations are deployed within the course of a single running text. In order to explore this 'textual' factor, we examine a selection of examples from specific datasets and corpora, including different genres and registers (although we concentrate here specifically on examples from scientific research articles in contrast with a science forum and a Wikipedia page with references to the titles of scientific research articles). On the basis of these examples, we attempt to answer a series of research questions:

1. What is the nature of the transitivity relation (i.e. the semantic role of Participant and Process) that holds between N1 and N2 in <N1 N2> vs. <N2 of N1>?
2. Is there a general preference for one structure over another ('packed' or 'unpacked') in different text sections of the same text (titles vs. abstracts vs. methods, etc.) or in a given sub-section of a specific text type (for instance in scientific article titles)?
3. What linguistic factors constrain the choice of one structure over another?

4. Is there a preference in the direction of variation (from 'packed' to 'unpacked' or from 'unpacked' to 'packed' and eventually back to the initial form) in a running text (as sentences run on into each other to form a linear text)?

We answer these four research questions in the following case studies. On the basis of our observations, we also propose a more general hypothesis on the textual or discourse function of the <N1 N2> vs. <N2 of N1> alternation in English, and on the relevance of such an alternation for the rhetorical and cognitive construction of scientific discourse.

In the first part of the paper, we give a brief overview on how the issues of alternation and variation have been conceptualised in the fields of general linguistics and ESP/LSP studies, especially in the study of terminology and phraseology (lexical collocations). In the second part of this article, we discuss the methodological aspects of our study on binominal alternation. We then select a sample of four alternations for a descriptive case study in the third section. In the conclusion, we discuss the impact of our analysis on our understanding of this dynamic lexico-grammatical pattern in science writing.

2 Current research on variation and reformulation

In this paper, we use the term *alternation* to refer to the switch between the alternative structures <N1 N2> and <N2 of N1>. To be more precise, we use the term *alternation* to refer to two processes which cannot be fully separated: firstly alternation in the usual sense of 'successive occurrence' of two different structures and secondly in the sense of a 'choice or alternative' between two different structures. However, before exploring this notion, we must examine how similar notions such as *variation* and *reformulation* have been conceptualised in various related areas of linguistics. While for some linguists, the term *variation* refers to how formal features of language vary according to function, text type or register (Biber 1988), for others, notably terminologists and phraseologists, the term *variation* has been associated with competing forms which are used for the same term or sequence of words within a given phrase. In this section, we examine what is meant by *variation* in those areas most concerned with the subject: text linguistics, terminology and collocation studies.

2.1 Variation and reformulation in text linguistics

The issue of alternation between pre-modified versus post-modified nominals can be studied from the point of view of formal syntax, especially in terms of equivalent transformational pairs. However, this approach, which does not take into account such notions as context and discourse function, is not the focus of our study. Instead, the notions of variation and reformulation, recently studied in text linguistics and discourse analysis, are more relevant here.

The notion of reformulation is a core feature of the functional approach to language. In particular, it can be traced back to the Prague School's Functional Sentence Perspective established by Firbas (1992), following Mathesius (1928) who first described this approach; see also Halliday 1999 (122–126), Halliday and Hasan's (1976) analysis of thematic progression and cohesion within texts, and Sinclair's model of 'posture' (Sinclair 1980, 1993, Sinclair/Mauranen 2006). According to these approaches, linguistic features are identified either as pre-cursors or (more often) as encapsulations of previous statements. Within the on-going text, meaning is thus seen to be gradually built, de-constructed, re-negotiated and, ultimately, kept in a

constant state of maintenance. Following Francis (1986, 1994), discourse analysts have thus focused on the role of specific items such as 'anaphoric nouns', 'labelling nouns' and so on as key features by which complex chains of lexical links are formed and re-formed throughout a text (Hoey 1991, Yamasaki 2008, Pic/Furmaniak/Hugou 2013, Bordet 2015). In one such study Gledhill (1995, 1999) looked at how the functions of 'prospection' and 'encapsulation' are realised in a sample of scientific research articles. Thus in one research article (Gledhill 1999: 7), the central chemical process described in the paper is reformulated first as a nominal expression of a methodology (*synthesis of …*), then as a strategy (*route, strategy*), and finally as various degrees of scientific claim (*total synthesis, methodology, general and versatile strategy, efficient synthesis* etc.):

(2a) **Synthesis** of the seleno compound [...]
(2b) This **route** provides d4T in six steps [...]
(2c) The completion of this **total synthesis** [...] establishes this **methodology** as a **general and versatile strategy** towards the **efficient synthesis** of a range of important antiviral nucleosides [...]

As we see in later sections, from the point of view of discourse analysis, the alternation between <N1 N2> and <N2 of N1> can be seen essentially as a choice between 'packing' a referent away as part of the background information in relation to the rest of the text (as in the structure <N1 N2>) or 'unpacking' a referent in order to place focus on it as a contrastive or 'new' piece of information (as in the structure <N2 of N1>).

2.2 Variation and reformulation in terminology studies

For a long time, variation remained out of the spotlight in works on terminology. The reason behind this is quite simple: the focus in traditional terminology was on 'stability', the key criterion for a term in order to fulfil its naming function in specialised discourse, in order to avoid ambiguity. In those cases in which reformulation created instability, variation was often perceived as an anomaly, and thus as an exception to the required stability of specialised lexicons (Pecman 2012: 29–31). However, this tendency has been progressively abandoned by many terminologists, who have begun to examine the phenomenon of terminological variation in terms of its communicative purposes (Faulstich 2000, Temmerman 2000, Tercedor 2011, Pecman 2012, 2014, Humbley 2015).

In terminology studies, variation is primarily considered in parallel with the phenomenon of synonymy ('competing terms'), and is often regarded as a sign of 'neonymy', because the process of coining new words generally gives rise to various denominations for the same concept. This approach has now been complemented by studies on the dynamic behaviour of terms within texts, as well as increased interest in the role of terminological variation in knowledge construction. As nominals have the capacity to name entities, it is well known that they are the most salient features in which various modifications or alterations appear to have specific functions. Following from this observation, variation is hence seen as a purposeful process, and not a mere sign of instability.

Tercedor (2011), for instance, studies terminological variation as a result of multidimensionality, that is: language users are free to choose the focus they want to place on an object/entity when naming it. Hence the term *rubber boots* focuses on perception, while the item

rain boots focuses on function. These are cognitively motivated concurrent ways of naming. Perceptual vs. functional properties, as in *rubber boots/rain boots*, are just one in a series of alternating cognitively motivated distinctions (others can include: location vs. function or material vs. shape) for different ways of seeing and thus naming.

Another type of variation, more in line with the object of our study, involves lexico-grammatical reconstruction or 're-packing'. This form of variation has been analysed by Ormrod (2003, 2004), whose approach (following Halliday 1998a, 1998b) sheds interesting light on how terms are created dynamically within on-going texts. In a comparative analysis of academic research articles written by native and non-native speakers of English, Ormrod shows how concepts acquire additional properties through term variation and modification, and how they function within the on-going text by building up knowledge through the construction of increasingly complex nominal groups:

> In scientific texts, it is the nominal group that has the greatest number of possibilities for semantic and syntactic realization. The scientific "object" is named – the Head word – and this Head gradually acquires supplementary properties as the experimental work unfolds, in the form of Modifiers and Qualifiers. This additional information can then be integrated into the Head in the form of complex or compound terms, eventually becoming lexicalised items. (Ormrod 2004: 51)

In previous studies on apparent tentativeness in naming scientific concepts, Pecman (2012, 2014) has demonstrated the specific cognitive and rhetorical functions behind variational paradigms with a high degree of terminological variation. She shows two types of motivation for maintaining variation in scientific texts: 'rhetorical', by which the scientists manage to place the focus on the novelty of their research, and 'cognitive' which serves to explicitly render and shape the meaning of a concept by switching from the packed to the unpacked form of a term.

The notions of *unfolding* (Ormrod 2004) and *unpacking* (Pecman 2014) follow on from the work of Halliday (1995, d. f. Halliday/Matthiessen 2014), who associate these processes with the more general textual function of 'grammatical metaphor':

> [Grammatical metaphor is] a process of reformulating a concept through different grammatical categories in order to capture its essence and establish a new paradigm of knowledge. As such, grammatical metaphor would be the central feature of scientific discourse and the pathway that leads from experience to theory, and consequently to new knowledge. (Pecman 2014: 2)

Finally, in a recent study by Mestivier-Volanschi (2015: 6 f.) on adjectival and participial compound pre-modifiers (e. g. *invasion-inhibitory assays* vs. *inhibition of invasion assays*), we find further proof of the importance of (un)packing processes in the construction of scientific discourse and knowledge. The author observes two types of constructions: advance and retrospective constructions (and which correspond to, respectively, unpacking and packing processes).

2.3 Variation and reformulation in collocational studies

We have so far examined reformulation and variation in relation to text linguistics and terminology studies, but there has also been a long tradition of work on this subject from the point

of view of phraseology and corpus linguistics. Of particular note are studies on collocational networks derived from scientific texts (Williams 1998) and collocational variation within scientific texts (Bordet 2013, 2015, Giacomini 2015). In this respect, it is worth noting a number of findings that are useful for understanding the notion of *variation*, and our use of this term.

For our purposes, it is sufficient to define phraseological units (or lexical collocations) as sequences of words which co-occur and are co-selected frequently within texts, and which have a degree of structural and semantic predictability (following Firth's [1957] principle that a meaning of a word depends on the 'company it keeps'). From a phraseological point of view, alternations such as <N1 N2> vs. <N2 of N1> can be seen as two distinctly specialised multiword units, rather than variations of the same structure. This approach to language is of course quite opposed to the transformational idea of a true alternation between two equivalent or concurrent forms. Furthermore, corpus-based work on collocations in ESP and LSP has become complementary to current thinking in terminology studies. The examples studied in each discipline are often very similar, as we can see for instance in Gledhill's (2011) analysis of the productive collocational relationship between *gene* and *express,* as illustrated in 3a and 3b:[3]

(3a) Under these conditions, we did not detect PAF-R **gene expression** (Ma and Bazan, 2000).

(3b) However, **expression of the gene** was not confined to the hair follicle, as the transgene phenotype included not only hair abnormalities, but also vertebral defects and bladder, liver and intestinal tumors.

For the phraseologist, such examples demonstrate that there is an underlying selectional restriction that operates not only at the level of the lexical items (here the Participant *gene* and the Process *express*)[4] but also within the extended co-text (in examples 3a–b, we are again dealing with the opposition <N1 N2> vs. <N2 of N1>). Thus, the point is not whether the two items form a single or a compound terminological unit, but rather how these forms not only co-occur within a certain span of words in proximity to each other, but also within a certain type of construction and within a certain position within the text. As mentioned above, this kind of observation has led text linguists to examine linguistic reformulation in terms of lexical chains within the on-going text (Bordet 2013, 2015).

In sum, the issue of intra-textual variation of expression, whether in text linguistics, terminology or collocation studies, has been widely investigated from various perspectives. However, we would suggest that a further perspective, namely the role of variation in the textual construction of meaning and knowledge, has remained much less systematically investigated, although such a notion has been alluded to in some studies (Halliday 1995, 1998b, 2003, Gledhill 1997, Pecman 2012, 2014). One of the reasons for the lack of research in this area may lie in the ways in which the dynamic features of texts have been examined up to now. We would suggest that many linguists, especially formal grammarians, pay considerable attention to phenomena such as grammatical paraphrase, but they look at alternation without taking into account any textual context. In the following sections, we attempt to correct this tendency

[3] Examples taken from the 500,000 word Pharmaceutical Sciences Corpus (PSC), reported in Gledhill (1995, 1997).

[4] Note that, by convention in the Systemic Functional model (Halliday 1985, Halliday/Matthiessen 2014), capitals are used for any term that refers to a clause function (Subject – Predicate, Head – Modifier, Process – Participant, etc.).

by looking at the discourse function of alternation within running text. We then examine how such variations as <N1 N2> vs. <N2 of N1> contribute to knowledge construction.

3 Context of the study and methodological aspects of investigating structural variation

In the following analysis of <N1 N2> vs. <N2 of N1> nominal alternations, we have adopted a multiple case study based on a selection of relevant examples taken from the scientific literature and science forums. Some of these examples were used in our Master's programme in specialised translation[5] at the Department for Applied Languages and Intercultural Studies – *Études Interculturelles de Langues Appliquées* (EILA)[6] – Paris Diderot University. On this course, students are taught how to compile and interrogate comparable corpora on various areas of specialised knowledge, in order to find solutions to various difficulties encountered during translation. They are given the opportunity to work on different domains and LSPs, including scientific discourse, and they discover a variety of tools for working with corpora (such as AntConc[7], SketchEngine[8] and IMS Open Corpus Workbench[9]). They are also taught how to conduct terminological analysis and to create lexical resources within an online term base – the ARTES[10] terminological and phraseological database – for the purposes of translation. ARTES is a mutidomain multilingual language resource and an experimental pedagogical tool intended to provide language resources for translators and scientists to help them write in their second language. We also see ARTES as a tool which helps our Masters students to acquire knowledge in terminological analysis and terminology management. The ARTES database offers students a template for selecting and organising knowledge-rich information of particular interest for specialised translators. In addition, the use of corpora by translation students and the ARTES project represent a rich 'exploration site' for our research unit in corpus linguistics for specialised translation, electronic lexicography and lexical resources design (Pecman/Kübler 2011, Kübler/Pecman 2012, Pecman 2012, 2014, Mestivier-Volanschi 2015, Kübler et al. 2016, etc.).

We would claim that the nominal alternation <N1 N2> vs. <N2 of N1> frequently occurs in English scientific discourse and represents a double source of difficulty for our students, whose native language is usually French. In the first instance, this alternation represents a problem in the process of translation, and it is worth stating again here that this alternation is particular to English and not found in the basic structure of French nominal groups. Secondly, the <N1 N2> vs. <N2 of N1> alternation represents a problem for terminological analysis, since our students often hesitate between one form or another when they need to determine the main entry term in the ARTES term bank. The problem is compounded by the fact that the students also have to consider whether one form or the other is an alternative form (a true synonym) or a context-bound variant.

---

5 Master's web page: http://www.eila.univ-paris-diderot.fr/enseignement/lea/master

6 EILA's web page: http://www.eila.univ-paris-diderot.fr

7 AntConc's web page: http://www.laurenceanthony.net/software/antconc

8 SketchEngine's web page: https://www.sketchengine.co.uk

9 IMS Open CWB's web page: http://cwb.sourceforge.net

10 ARTES stands for *Aide à la Redaction de TExtes Scientifiques*/Dictionary-assisted writing tool for scientific communication. Database interface: https://artes.eila.univ-paris-diderot.fr and ARTES project web pages: http://www.eila.univ-paris-diderot.fr/artes

In order to examine the role of the <N1 N2> vs. <N2 of N1> in scientific discourse and the parameters that govern the choice of one form over another, we have selected a sample of alternations which will serve as four case studies, namely:

- ***synthesis of aspirin*** vs. ***aspirin synthesis***[11],
- ***hurricane development*** vs. ***development of hurricane***[12],
- ***hydrothermal plume release*** vs. ***release of a hydrothermal plume***[13] and
- ***oligonucleotide synthesis*** vs. ***synthesis of oligonucleotide(s)***[14].

All of these examples were selected from two scientific domains – Earth and planetary sciences (EPS) and chemistry. The first example (*synthesis of aspirin …*) illustrates an example of online scientific discourse, while the following three involve examples retrieved from scientific research articles, allowing us to consider two varieties of scientific communication: an academic published article and online discussion. The first and the fourth examples (*synthesis of aspirin* and *oligonucleotide synthesis*) were retrieved from online sources and have been used as a pedagogical material to introduce the binominal alternation to our students (as mentioned above). The second and the third examples (*hurricane development* and *hydrothermal plume release*) illustrate the problems encountered in a corpus on Earth and planetary sciences compiled by Master's students for conducting terminological analysis and resolving translational difficulties (as mentioned above). Furthermore, the first three examples have already been reported in our previous research (Gledhill 1995, 1999, Pecman 2102, 2014). However, the focus of that previous research was not on the systematic analysis of <N1 N2> vs. <N2 of N1> nominal alternations, the main focus of the present paper. (The source for each example is further described in the following sections, in which each example is analysed in turn.)

As our aim is to examine both the role of the alternation <N1 N2> vs. <N2 of N1> and the parameters that govern the choice of one form over another, each example is investigated on the basis of distributional analysis, presuming that any two structures having different meanings, values or functions also differ somewhere in their distribution, that is the contextual environments in which they appear (Harris 1951, 1954). Distributional analysis offers a means for observing both the patterns of alternation across a text, and contextual tendencies that constrain the choice of one structure over another.

Our focus here is on the alternation between 'agnate forms' (Halliday's term for forms which are related lexico-grammatically, but not necessarily derived from each other by any underlying transformation) which occur in the same text ('intratextual variation') or which occur in the same section of a specific text type. We also limit our search to nominal constructions (for instance *protein synthesis*, *synthesis of proteins*) involving a transitive semantic relation be-

---

11 Source: The Science Forum 'The Naked Scientists' based at Cambridge University's Institute of Continuing Education, accessed 11 July 2015: <http://www.thenakedscientists.com/forum/index.php?topic=27009.0>.

12 Source: EPS corpus (Pecman 2012, 2014), Arpe, K./Leroy, S.A.G. (2009): "Atlantic hurricanes – Testing impacts of local SSTs, ENSO, stratospheric QBO – Implications for global warming." *Quaternary International* 195: 4–14.

13 Source: EPS corpus (Pecman 2012, 2014), Dziak, R. P./Bohnenstiehl, D. R./Cowen, J. P./Baker, E. T./Rubin, K. H./Haxel, J. H./Fowler, M. J. (2007): "Rapid dike emplacement leads to eruptions and hydrothermal plume release during seafloor spreading events." *Geology* 35.7: 579–582.

14 Source: Wikipedia article on Oligonucleotide synthesis accessed 20 October 2015: <https://en.wikipedia.org/wiki/oligonucleotide_synthesis>.

tween a Material process and a Medium participant[15]. Thus, for example, in 'protein synthesis' *synthesis* is a Material process and *protein* is the Medium for that process (that is the result, product or affected object which is associated with that process). In the following section we look at constructions and patterns which satisfy these criteria.

Furthermore, our analysis of the first three cases focuses on determining the patterns of alternation and the contextual and textual constraints which condition the alternation, while the last case complements these analyses by providing a detailed study of the contextual factors involved in determining the choice of one structure as opposed to the other.

4 Analysis of binominal structural alternations in context

4.1 Aspirin synthesis vs. synthesis of aspirin

This type of alternation was first examined in Gledhill (1995, 1999). In that study, the textual development of a series of complex nominals was analysed in a sample of ten research articles in the biochemical and medical sciences. In many of these texts, chains of reference were built up around an item such as *synthesis* (a term which turns out to be a key process in this field). We would suggest that it is no accident that a Google search throws up many examples in which the sequence 'synthesis of N' co-occurs in close proximity with the equivalent 'N + synthesis'. The example we cited above consists of three lines (1a–c) and was found within the first few search results:

(1a) What is the **aspirin synthesis** mechanism?
(1b) Essentially i'm looking for a proper curly arrow type mechanism for the **synthesis of aspirin** from salicylic acid (2-hydroxybenzoic acid) and ethanoic anhydride.
(1c) Curiosity strikes ... what *do* you want a mechanism for **aspirin synthesis** for, then??

This is a significant example because it shows that: i) an alternation of this type can occur even in a context of an online exchange between self-selected experts and other 'interested parties', such as found on forums, where the style is not as constrained as in academic research articles, and ii) there is a perceptible difference of rhetorical function in which the alternation occurs. We would suggest that this change in function has an impact on the selected forms and their alternation (i.e. packed <N1 N2>, then unpacked <N2 of N1>, then back to packed again <N1 N2>).

Let us now look at each stage of the exchange (numbered 1a to 1c) in turn. In the first part of the exchange (1a), the direct question (which also doubles as a title) involves the <N1 N2> form as a pre-modifier within a longer nominal group: ***aspirin synthesis*** *mechanism*. We can assume here that the main focus of the question is the final element *mechanism*. According to Halliday's (1995, also Halliday/Matthiessen 2014: 114–121) analysis of thematic structure, the final position in a clause typically corresponds to an 'unmarked focal position'. The packed <N1 N2> form is thus being used here in place of two alternative structures (the other possibilities being: *what is the mechanism for the* ***synthesis of aspirin*** */ what is the mechanism for* ***aspirin synthesis***). In both cases, the process (*synthesis*) and its result (*aspirin*) would be in focal position, but in the original text, the speaker/writer has decided to avoid this. By doing this, the

[15] The terminology we use for semantic roles is adapted from Systemic Functional Grammar (Halliday 1985, Halliday/Matthiessen 2014).

author is implicitly signalling that *aspirin synthesis* is a 'given' piece of information which is not 'at stake' or 'up for grabs' at this stage of the text.

In the second stage of the exchange (1b), the 'original poster' (OP) is offering an explanation for his/her question. The nominal group (NG) here is still headed by the noun *mechanism*, but this is now pre-modified by much more complex structure. Indeed this structure involves two recursively embedded NGs as pre-modifiers, which can be bracketed as: *a proper [[curly arrow] type]* ***mechanism***. Not only is our NG heavily pre-modified, it also involves complex post-modification: *synthesis [of aspirin [from salicylic acid (2-hydroxybenzoic acid) and ethanoic anhydrid]]*. Thus the unpacked NG *synthesis of aspirin* is itself embedded within two prepositional phrases, the second of which specifies the source of the *aspirin* as a product, as in ***for*** *the synthesis of aspirin* ***from***. Although it is possible to replace the unpacked form *synthesis of aspirin* with the packed form *for aspirin synthesis from,* it occurs to us that the author of this passage is more or less obliged to achieve some kind of structural balance in this very complex phrase. In other words, the choice of the unpacked form here has to do as much as with informational content (or focus) as with the extent to which the structure of the NG involves complex pre- and post-modification. As we discuss below, we call this kind of structure 'embedding', and in the final section of our analysis (devoted to the study of *oligonucleotide synthesis* vs. *synthesis of oligonucleotide(s)*) we examine to what extent this type of constraint affects which structure is actually chosen in a given context.

In the third and final part of this exchange (1c), a new contributor intervenes and uses the packed <N1 N2> alternation. Unlike the first part of the exchange, it is notable that the focus here is no longer on the final element in the clause: instead, the focus is on the polarity of the question itself, which is marked typographically and syntactically by the auxiliary *do*. As mentioned above, unmarked focus typically falls on the final element in the clause. However, marked focus can in theory be placed in any position, and this is often signalled by using some prosodic or typographic cue. We would suggest that the author of this passage is using the packed <N1 N2> form here largely because the process (*synthesis*) and its product (*aspirin*) are not 'at stake' at this point in the exchange, and are expressed once again as 'given' (or 'known' information), as in the first sequence. This observation allows us to posit a rough hypothesis at this stage: the use of a packed NG allows for the expression of new information elsewhere in the co-text, whereas the use of an unpacked NG suggests that the NG itself carries a degree of (contrastive or new) focus.

4.2 Hurricane development vs. development of hurricane

This type of alternation was first reported in Pecman (2012, 2014) where it was identified within a corpus of scientific articles on Earth and planetary sciences (EPS) which were sectioned according to the general structure of scientific articles (IMRAD), allowing for the observation of the distribution of variants across these sections. This example offers proof that binominal structural alternations have a particular directionality within a text, namely from packed to unpacked and back to packed (cf. figure 1).

A further hypothesis that one might want to make is that the packed version of the NG, *hurricane development*, is 'term-like': in other words given a potential alternation <N1 N2> vs. <N2 of N1>, it is the packed structure that tends to be the preferred form for technical terms in English. In this example, the packed form appears 23 times throughout the text, while the only two occurrences of the unpacked structure, *development of hurricane*, appear in the middle

part of the article, in the section with the subtitle: *Other impacts on **hurricane development** than the in-situ SST*. (The <N1 N2> form thus appears again in the title.)

As figure 1 shows, we find occurrences where both variants appear in very similar contexts, suggesting that <N1 N2> and <N2 of N1> are free concurrent structures: *(important) (factor) for hurricane development/the development of hurricane* (with hits 6 and 15 for packed and 9 and 10 for unpacked form):

Figure 1: Alternating pair hurricane development and development of hurricane observed in the same text

text_id	Left context	Keyword	Right context
1_abstract_CLIM_2009_Arpe	the atmospheric stratification Processes affecting	hurricane development	over the North Atlantic like
1_abstract_CLIM_2009_Arpe	trend. The area of	hurricane development	is limited among others by
4–6_part1of6_CLIM_2009_Arpe	The main processes leading to	hurricane development	The main features of tropical
4–6_part1of6_CLIM_2009_Arpe	circulation is needed for initiating	hurricane development	. This frequently stems from
4–6_part4of6_CLIM_2009_Arpe	5. Other impacts on	hurricane development	than the in – situ
4–6_part4of6_CLIM_2009_Arpe	signals. Important for the	hurricane development	over the Atlantic is an
4–6_part4of6_CLIM_2009_Arpe	averaged for the area of	hurricane development	are shown. Means of
4–6_part4of6_CLIM_2009_Arpe	leading to a suppression of	hurricane development	. Fig. 4 demonstrates
4–6_part4of6_CLIM_2009_Arpe	/ dp) for the	development of hurricanes	by comparing time – series
4–6_part4of6_CLIM_2009_Arpe	Another important factor for the	development of hurricanes	is the existence of a
4–6_part4of6_CLIM_2009_Arpe	shear in the area of	hurricane development	as shown in Fig.
4–6_part4of6_CLIM_2009_Arpe	phase one can expect more	hurricane development	and during the eastern phase
4–6_part4of6_CLIM_2009_Arpe	of hurricane occurrences. The	hurricane development	area is bordered in the
4–6_part4of6_CLIM_2009_Arpe	5 1C that limits the	hurricane developments	to a small belt between
4–6_part6of6_CLIM_2009_Arpe	, that are important for	hurricane development	, will be investigated under
4–6_part6of6_CLIM_2009_Arpe	suggests that the area of	hurricane development	might increase as well.
4–6_part6of6_CLIM_2009_Arpe	But the unchanged area of	hurricane development	could also be due to
4–6_part6of6_CLIM_2009_Arpe	for the area of possible	hurricane development	as simulated by the model
4–6_part6of6_CLIM_2009_Arpe	changes in the area of	hurricane development	between 1961 , Âì1990 and
4–6_part6of6_CLIM_2009_Arpe	only this phase which suppresses	hurricane developments	. Table 1 gives an
4–6_part6of6_CLIM_2009_Arpe	impacts of different parameters on	hurricane development	in a future climate under
4–6_part6of6_CLIM_2009_Arpe	An in – crease of	hurricane development	in the future can be
7_conclusion_CLIM_2009_Arpe	, ENSO and QBO on	hurricane development	have been demonstrated. Warmer
7_conclusion_CLIM_2009_Arpe	the latter two affect the	hurricane developments	through the modification of the
7_conclusion_CLIM_2009_Arpe	scale quantities, to which	hurricane developments	are sensitive, have been
7_conclusion_CLIM_2009_Arpe	Atlantic. The area of	hurricane development	is limited among others by

Grafik nur 150 dpi

Let us look at the textual context in which the alternation takes place and determine the parameters that govern the alternation:

> [SUBTITLE]
> Other impacts on **hurricane development** than the in-situ SST
>
> [BODY OF THE TEXT]
> […] In Fig. 3a, vertical profiles of the zonal wind (U) for the season June–November averaged for the area of **hurricane development** are shown. Means of El Nino event years are compared with overall long-term means. There is a stronger vertical wind shear during El Nino events leading to a suppression of **hurricane development**. Fig. 4 demonstrates the importance of vertical wind shear (du/dp) for the **development of hurricanes** by comparing time-series of major hurricane frequencies with the difference of zonal winds at 200 minus 850 hPa for the area 5–151N, 50–701W, which is a measure for the vertical wind shear. […] Another important factor for the **development of hurricanes** is the existence of a moist-unstable stratified atmosphere. […] The QBO has an impact on the vertical wind shear in the area of **hurricane development** as shown in Fig. 3a.

A number of structural constraints can be observed here. Most notably, it would be difficult to envisage an unpacked structure in the first two <N1 N2> examples (which would become

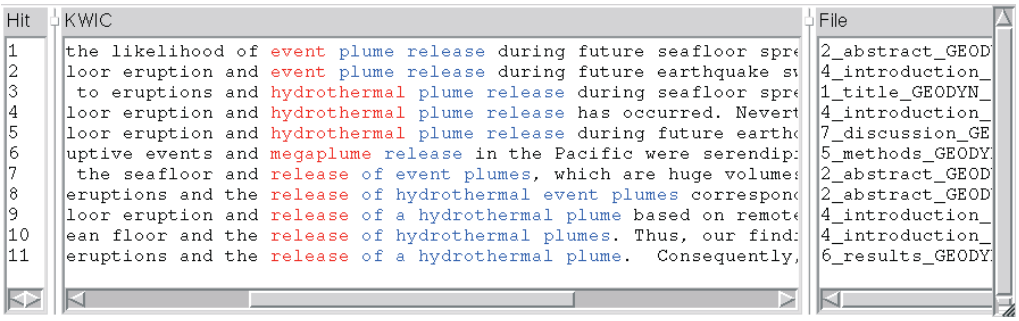
bottom-heavy, as in *the area of ?the development of hurricanes* or *a suppression of ?the development of hurricanes*). Conversely, for the following two examples of <N2 of N1>, it is entirely possible to have <N1 N2> (*the importance of wind-shear for hurricane development,* and *another important factor for hurricane development*). More generally, we might hypothesise that the <N2 of N1> structure may be blocked when the NG is a complement of a predicative N involving the preposition *of* (as in *the area of* and *a suppression of*), but not when it is introduced by a 'facette noun' involving another preposition (a facette noun is a complex pre-modifying NG which is the syntactic head but not the semantic head of the larger NG as a whole, as in ***an important factor for*** *the development of hurricanes*). The <N2 of N1> structure also does not appear to be blocked when the NG as a whole constitutes a circumstantial prepositional phrase which is not embedded in another NG (as in *for the development of*). Thus there seems to be a complex interplay between conceptual development (from a packed form to an unpacked form and then back again), as well as syntactic constraints governing the use of <N1 N2> at the beginning and the end of the text.

4.3 Hydrothermal plume release vs. release of a hydrothermal plume

This example was analysed in detail in Pecman (2012). We mention this example here to illustrate that the alternation <N1 N2> vs. <N2 of N1> can be a part of an extended variational set. Figure 2 shows the variational paradigm for the same concept throughout the same text. In Pecman (2012) we conducted a distributional analysis of variants across different subsections of the text, and suggested that the purpose of variation here is to create a 'novelty effect', that is to build new knowledge and attract the reader's attention to newly created terms, and hence a concept, by switching focus between a new (unpacked) and given (packed) structure.

Figure 2: Alternating pairs hydrothermal plume release vs. release of a hydrothermal plume and event plume release vs. release of event plumes within an extended variational set observed in the same text (reproduced from Pecman 2012: 37)

Grafik nur 150 dpi



The textual contexts where the alternating pairs occur are found in the first sections of the text, which are as follows:

> [TITLE]
> Rapid dike emplacement leads to eruptions and **hydrothermal plume release** during seafloor spreading events

> [ABSTRACT]
> The creation of ocean crust by rapid injection of magma at mid-ocean ridges can lead to eruptions of lava onto the seafloor and **release of event plumes**, which are huge volumes of anomalously warm water enriched in reduced chemicals that rise up to 1 km above the seafloor. Here, we use seismic data to show that seafloor eruptions and the **release of hydrothermal event plumes** correspond to diking episodes with high injection velocities and rapid onset of magma emplacement within the rift zone. These attributes result from high excess magma pressure at the dike source, likely due to a new influx of melt from the mantle. These dynamic magmatic conditions can be detected remotely and may predict the likelihood of **event plume release** during future seafloor spreading events.
>
> [INTRODUCTION]
> In studies of the global mid-ocean-ridge system, one of the most important discoveries during the past 15 yr was the first real-time detection of a dike injection and eruption associated with a seafloor spreading event, and a multidisciplinary effort was rapidly assembled to investigate in situ the many biological, chemical, and hydrothermal aftereffects on the attendant seafloor and water-column ecosystems (Fox et al., 1995; Embley et al., 1995; Baker et al., 1995). The key to rapid in situ investigations is the accurate evaluation of the real-time, remotely detected seismic data indicating that a large-scale seafloor eruption and **hydrothermal plume release** has occurred. Nevertheless, after nearly two decades and the real-time detection of seven mid-ocean-ridge seafloor spreading events, evaluation of the likelihood of a seafloor eruption and **release of a hydrothermal plume** based on remote seismicity has had a success rate of only ~60 %.

Once again, we can see a correspondence here between general structural constraints within the English NG and the role of information structure within the individual text. As mentioned above, there appears to be a general structural restriction on NGs which blocks expansion to <N2 of N1> in the presence of complex pre-modifiers (which introduce the NG as a complement) or complex post-modifiers (determined by the NG in question). This can be seen at the end of the example above where *the likelihood of **event plume release** during future seafloor spreading events* cannot be comfortably reformulated as: *the likelihood of the ?**release** of **event plumes** during future seafloor spreading events*. In this example (and in the title), the restriction happens to coincide with the need to place the circumstantial (*during future seafloor spreading events*) in focal position at sentence-final position.

Alongside the formal features of this example of alternation, we also have to take into account the functional features of the text. As can be seen here, the unpacked form can be used in knowledge-rich contexts which provide useful information on the meaning of the concept, for instance by means of a relative clause (for instance ***release of event plumes**, which are huge volumes of anomalously warm water enriched in reduced chemicals*), while packed forms are used when additional information or knowledge is provided on the processes involving the concept referred to by the term (for instance *the likelihood of **event plume release** during future seafloor spreading events*).

Based on our first three examples, the formal switch from <N1 N2> to <N2 of N1> and back seems to occur for specific reasons, notably in order to adapt the structure of a com-

pound nominal according to the relative 'weight' of nominal expansions which modify the N1 and/or the N2. This leads us to believe that the binominal alternation may be a useful device for constructing meaning in scientific discourse. These examples also show that the <N1 N2> vs. <N2 of N1> alternation can appear as a single form of variation in a text or as a part of more extended variational set.

This type of fluctuation in the structure of nominals can be related to two similar, albeit distinct, functions: rhetorical and cognitive (Pecman 2012, 2014). The first consists in using variation for the purposes of presenting a concept alternatively as 'given' and thus indisputable or 'new' and thus open to debate. The given-new switch has the effect of attracting the reader's attention to the concept at stake. The second function consists in unfolding the term structure and rendering its meaning explicit by abandoning nominal structure and turning the information into a clause.

4.4 Oligonucleotide synthesis vs. synthesis of oligonucleotide(s)

The example in the previous section demonstrates the extent to which alternations such as *synthesis of aspirin* vs. *aspirin synthesis, hurricane development vs. development of hurricane* and *hydrothermal plume release* vs. *release of a hydrothermal plume* may be sensitive not only to textual function but also to structural context. It might be hypothesised that certain forms, such as the packed pattern, might correlate with certain parts of a text (such as title sections), or might be associated with certain structural contexts (such as heavy pre- or post-modification). In order to test this further, we searched for data containing a high concentration of both unpacked and packed NGs involving the same nominals within the same text section, a title. One of the sources providing such data is a Wikipedia entry on *oligonucleotide synthesis*.

This page contains a references section listing 121 research article titles, in which there are 73 instances of *oligonucleotide synthesis* or *synthesis of oligonucleotide(s).* We would suggest that this dataset is of particular interest, because it contains a high density of both structures produced within the same context (namely a reference to the same concept in a selection of related research article titles). In table 1 below, we set out the distribution of two structures according to the context type (note that, in the following analysis, we have accepted various lexical substitutes for *oligonucleotide*, such as: *oligoribonucleotide*, *DNA*, *RNA sequence,* etc.):

Table 1: Frequencies of packed and unpacked forms in Research Article Titles (from one Wikipedia page)

Context/Structure	Packed form (<N1 N2>): *oligonucleotide synthesis*	Unpacked form (<N2 of N1>): *synthesis of oligonucleotide(s)*
Context free structure	1	5
Bottom-heavy structure (heavily postmodified)	1	31
Top-heavy structure (heavily premodified)	22	13
Total examples (% out of 73)	24 (33 %)	49 (67 %)

These figures cannot be seen as exhaustive or representative of all research articles, but they do suggest that i) titles in scientific research papers do not make exclusive use of either the packed or the unpacked form, ii) there is however a certain preference for the unpacked form

in 'bottom-heavy' contexts, that is to say with heavily post-modified NGs, and in certain cases without contextual constraints, and iii) there is some 'free variation', between packed and unpacked forms in 'top-heavy' contexts. In order to interpret these figures meaningfully, we look at each of the following cases in turn: 1) context-free cases, 2) 'bottom-heavy' contexts, 3) 'top-heavy' contexts.

First, we examine examples of both structures with zero contextual constraints, that is examples where neither the first N nor the second N of the NG is pre-modified or post-modified, as in:

(4a) **<N1 N2> context free structure**:
Iyer, R. P.; Beaucage, S. L. 7.05. *Oligonucleotide Synthesis.* In: Comprehensive Natural Products Chemistry, Vol. 7: DNA and Aspects of Molecular Biology. Kool, Eric T.; Editor. Neth. (1999), Elsevier, Amsterdam, pp. 105–152.

(4b) **<N2 of N1> context free structure**:
Ogilvie, K. K.; Theriault, N.; Sadana, K. L. (1977). *"Synthesis of oligoribonucleotides".* J. Amer. Chem. Soc. 99 (23): 7741–7743. doi:10.1021/ja00465a073.

Generally, in cases of zero contextual constraints, the <N2 of N1> structure seems to be more frequent as shown in table 1 (although it has to be emphasised again that in corpus-linguistic terms, this is a very small sample).

Looking at 'bottom-heavy' structures (4c), which involve the modification of the second nominal, there is a clear preference in our data for unpacked, <N2 of N1>, forms. In such cases, the second N is often pre-modified as well as post-modified, and the post-modifier is often a defining reduced clause introduced by a verb expressing a relation of possession (*containing, carrying, having*).

(4c) **'bottom-heavy' <N2 of N1> structure**:
Synthesis of oligonucleotides carrying 5'-5' linkages using copper-catalyzed cycloaddition reactions.;
Synthesis of chimeric *oligonucleotides* containing phosphodiester, phosphorothioate, and phosphoramidate linkages.;
Synthesis of a dithymidine *dinucleotide* containing a 3': 5'-internucleotidic linkage.

However, it is interesting to note that when the whole NG is post-modified by a prepositional phrase (4d), the <N1 N2> structure is preferred. It may be the case that the authors have chosen the packed <N1 N2> form because the N1 is so short, although in this case the unpacked structure would have been just as grammatically well-formed. As we see below, when the N1 is *DNA* or *RNA*, there is a preference for the packed form. This suggests that <*DNA* N> or <*RNA* N> might be a 'lexicalised' or terminologically stable name for this process. In other words, the abbreviated form may have been selected as a preferred terminological unit (rather like a prefix) because it allows for packed form to be easily modified:

(4d) **'bottom-heavy' <N1 N2> structure**:
DNA synthesis without base protection.

Finally, there are two 'top-heavy' structural possibilities: i) heavy pre-modification of the first N (4e and 4f, see below), ii) the embedding of the NG within a (longer) pre-modifying structure (4g and 4h, set out below). So far we have seen that the packed/unpacked alternation is

often determined by the grammatical context in which the NG is used. However, for each of the 'top-heavy' patterns we have identified, there is also a phraseological correlation, that is to say that each of these structures also involves a degree of co-selection between the grammatical structure and an associated family of semantically-related lexical items. It appears that in each sub-pattern, there is also a preference for either the packed or the unpacked form.

In our first sub-pattern, which we call the 'top-heavy <N2 of N1> classifier' pattern, we find that an unpacked structure is preferred when the first nominal, N2, is classified by a pre-modifier and defined as a specific type of synthesis (*stepwise, electrochemically-directed*, *solid-support*), as in (4e). It is notable that the first nominal in these examples, the N2, is a specific discourse referent, but the product of the synthesis, expressed in the second nominal, the N1, is often more general when compared with the bottom-heavy pattern (example 4c): instead, here we see that the product is *oligonucleotide(s)* or a related item without further qualification:

(4e) **'top-heavy' structure, the top-heavy <N2 of N1> classifier pattern**:
Syringe method for stepwise chemical *synthesis of oligonucleotides*.;
Electrochemically-directed *synthesis of oligonucleotides* for DNA microarray fabrication.;
Solid support *synthesis of* all-Rp-**oligo(ribonucleoside** phosphorothioate)

In our second sub-pattern, the 'top-heavy <N2 of N1> embedded' pattern, we find that an unpacked structure is preferred when the NG as a whole is 'embedded', that is to say introduced by a subordinating preposition (signalled by *for* or *in*), which in turn relates the chemical process (*synthesis*) to the preceding item. It is notable that, in contrast to the zero constraint pattern (4a), which in structural terms is a similarly 'balanced' nominal group, the antecedent N in (4f) is a generic term for a chemical reagent or group of reagents (*groups, intermediates, reagents*), while the product of the synthesis (introduced by *of* N) is usually a very specific item (*3'(2')-O-amino-acylated RNA sequences, internucleotide phosphate analogs, two DNA 51-mers*):

(4f) **'top-heavy' structure, the top-heavy <N2 of N1> embedded pattern**:
Novel fluoride-labile nucleobase-protecting *groups* ***for*** the *synthesis of* 3'(2')-O-amino-acylated *RNA sequences*.;
Deoxynucleoside H-phosphonate diester *intermediates* ***in*** the *synthesis of* internucleotide phosphate *analogs*.;
Hindered dialkylamino nucleoside phosphite *reagents* ***in*** the *synthesis of* two DNA *51-mers*.

In contrast, in what we call the 'top-heavy <N1 N2> pre-modifier' pattern, the N1 is pre-modified, while the N2 is left unmodified (note that any following prepositions are modifiers of the NG as a whole, not of the N2 specifically, as in 4d). In each example, the NG as a whole is typically followed or post-modified by a phrase which specifies the methodology used in the chemical reaction (by use of various items and structures, but especially: *with*, *without, use of, using, via*):

(4g) **'top-heavy' structure, the top-heavy <N1 N2> pre-modifier pattern**:
H-phosphonate *DNA synthesis* **without** amino protection.;

Phosphorothioate *oligonucleotide synthesis* **via** phosphoramidite chemistry.;
Ultrafast cleavage and deprotection of *oligonucleotides synthesis* and **use of** CAc derivatives.

Finally, in the 'top-heavy <N1 N2> classifier' pattern, which is the dominant pattern for top-heavy structures, the N1 refers to a general concept (often expressed in just one lexical item, like *DNA*, *RNA* or *oligonucleotide*), while the N2 tends to be unmodified. Generally, the NG as a whole is embedded after a preposition such as *during* or *in.* In this configuration, we have a prepositional phrase that acts as a circumstantial modifier of 'location' or 'duration':

(4h)**'top-heavy' structure, the top-heavy <N1 N2> 'classifier' pattern**:
5-(Benzylmercapto)-1H-tetrazole as activator for 2'-O-TBDMS phosphoramidite building blocks **in** *RNA synthesis*. […];
New Product: 0.5M CSO for non-aqueous oxidation **in** *DNA synthesis*; Removal of t-butyldimethylsilyl protection **in** *RNA-synthesis*.;
Efficient activation of nucleoside phosphoramidites with 4,5-dicyanoimidazole **during** *oligonucleotide synthesis*.

In summary, the examples we have seen above suggest that structural balance does appear to play a very important role in determining whether one or another alternative form will be used. Furthermore, since different lexico-grammatical structures tend to be associated with different rhetorical functions, we can also observe some interesting correlations between the overall function of the NG and both of the alternative forms. For example, if we concentrate on just the 'top-heavy' examples, we can see that there is a preference for the unpacked <N2 of N1> structure when the whole phrase is used to single out either a specific type of *synthesis* or a specific referent for N2 (i.e. a reagent or product that is more specific than *oligonucleotide*). In general, this confirms the function of the unpacked form we observed in our previous examples (*synthesis of aspirin* vs. *aspirin synthesis, hurricane development* vs. *development of hurricane* and *hydrothermal plume release* vs. *release of a hydrothermal plume*): the <N2 of N1> structure allows one of the nominals to be singled out as an available discourse referent and to be interpreted as being 'at stake'. On the other hand, as we have seen above, the <N1 N2> pattern is also used in top-heavy contexts, but especially in contexts where the NG as a whole is either used as part of a circumstantial modifier, or is used to refer to a more or less lexicalised process such as *DNA synthesis* or *RNA synthesis*. Once again, this suggests more generally that the <N1 N2> pattern is used to refer to a product (or Medium in Systemic Functional terms) of a process which is 'given' and thus can be presented in a packed structure.

5 Conclusions

Our analysis of a relatively small selection of nominal alternations has shown that equivalent packed and unpacked nominal groups represent a particularly rich resource for creating textual variation in English scientific discourse. It might be thought that there is a subtle semantic difference between these alternate forms. For example, it might be thought that the semantic relation between the N1 which expresses a product *(aspirin, hurricane, hydrothermal plume)* of a process (*synthesis, development, release*) appears to be neutralised or hidden in the packed form of the nominal (<N1 N2>, e.g. *aspirin synthesis*), whereas it might be thought that this process is more transparent in the unpacked form (<N2 of N1>, e.g. *synthesis of aspirin*).

However, the analysis we have presented above, despite the small amount of data, suggests that no such one-to-one relationship between syntax and semantics exists. Instead, the examples analysed in this study suggest that both the overall textual context as well as the local grammatical co-text, appear to play an important role in determining whether one or another structure will be preferred. In addition, as seen in section 4.4 above, it is difficult to quantify precisely which specific structural constraints determine the choice of packed or unpacked form in every example.

However, we can establish a set of hypotheses on the basis of our observations. First, we suggest that there is a general tendency to use the packed form in titles (in two-thirds of cases), while in the body of the text (whether in the following sentence, paragraph or sections) this form is often reformulated by the unpacked <N2 of N1> form. Thus, there appears to be a preferred direction of 'packing', which goes from 'packed' <N1 N2> towards 'unpacked' <N2 of N1> and then back to <N1 N2> (as in *hydrothermal plume release* switching to *release of hydrothermal plume* and back to *hydrothermal plume release*). As Pecman (2012, 2014) has observed, this represents the author's need to repeatedly switch focus from a concept being viewed as 'given' to being perceived as 'new', as other information is brought into play in the on-going argumentation of the text. As mentioned in our analysis above, the unpacked form generally announces that the concept is discussed or to be viewed as problematic and open to debate, while the packed forms are presented as a 'given' piece of information, which is 'up for grabs' at certain stages of the text. Consequently, the process of packing and unpacking contributes to the thematic progression of the text and its overall cohesion, and by a shift of focus allows speakers and writers to put the emphasis on novelty/innovation as well as to render some specialised concepts explicit.

It is however interesting to observe just how much variation there can be, even with research article titles in a particular domain. Thus the unpacked form <N2 of N1> (as in ***Synthesis of*** *chimeric* ***oligonucleotides*** *containing phosphodiester*) is preferred, because it is bottom-heavy. On the other hand, we encounter <N1 N2> in examples such as *Efficient activation of nucleoside phosphoramidites with 4,5-dicyanoimidazole during* ***oligonucleotide synthesis*** because the unpacked alternation would presumably attract too much focus here. It is notable that in each of these examples, there is clearly a phraseological constraint, but this does not relate to fixed sequences of items; rather there appears to be a degree of conventional co-selection which related one grammatical structure and grammatical context (e.g. N2 of N1 'is co-selected with' a post-modifying clause expressing possession, N1 N2 'is co-selected with' heavy pre-modifying specifiers and group-final position, etc.).

Furthermore, we have observed here and in previous studies that when a particular pair of lexical items undergo a cycle of reformulations and alternations, this appears to signal that they constitute key notions within that discourse, notions that are at the core of scientific demonstration. It is also the case that when an alternation occurs in the text, it often corresponds to a specific rhetorical function (a point attributed to Hoey 1991, and recently explored by Bordet 2013, 2015 in relation to chains of lexical reference). Thus, unpacked forms are often used in a specific type of knowledge-rich context exploited in terminological studies as 'defining contexts' (as in the example: ***release of event plumes****, which are huge volumes of anomalously warm water enriched in reduced chemicals*), while packed forms in the same texts are used when additional information or knowledge is provided (for instance: *the likelihood of* ***event plume release*** *during future seafloor spreading events*).

We have begun here to explore a distributional model for nominal group formulation in scientific English. This model exploits contextual constraints, both formal and functional. Our results suggest that the mechanism of alternation is not purely a question of style or 'open choice', but is also highly relevant to the construction of meaning. Our initial findings also suggest that the <N1 N2> / <N2 of N1> alternation is a key part of the underlying competence of any native (or 'native-like') user of English, especially at this very advanced and specialised level of proficiency (English for Academic / Scientific Purposes). It is also clear from the examples that we have explored above that alternations in nominal structures appear to be a specific language mechanism in English that participates in the construction of discourse by fulfilling different roles: cohesive (textual), rhetorical (interpersonal) and cognitive (experiential).

The hesitations that we first observed in French learners' translations into English when choosing between <N1 N2> and <N2 of N1> are now easier to understand, as we have demonstrated the complexity of establishing a distributional model (all the more so as this type of alternation does not exist in the learners' usual language, French), and of the need to understand better the alternation processes in the context of translation studies.

As the observations we have set out in this paper are only based on the study of a small sample of striking examples, further studies on alternating structures will certainly be necessary. We specifically intend to conduct a more systematic analysis of the <N1 N2> / <N2 of N1> alternation using representative corpora. A corpus-based analysis should help us to examine questions such as how alternations of this type operate in the general language, and how they function within different types of specialised discourse. Questions such as these may prove to be especially rewarding, as it may turn out that the alternation <N1 N2> / <N2 of N1> has evolved as a very specific feature of scientific and technical English over many centuries. If that is the case, and assuming that there are many other Process-Participant alternations of this type, what was their initial distribution in the language, and to what extent have alternations of this type undergone diffusion from structure to another and one specialised discourse to another? To the best of our knowledge, such questions still remain to be explored.

References

Biber, Douglas (1988): *Variation across speech and writing.* Cambridge: Cambridge University Press.

Bordet, Geneviève (2013): "The Rhetorical Role of 'Collocational Chains' and their Implications in the Building of Scientific Discourse Strategies." *European Journal of English Studies* 17.3: 235–250.

Bordet, Geneviève (2015): "Labeling Discourse to Build Academic Persona." Journal of Academic Writing 5.1: 106–118. 27.11.2017 <http://e-learning.coventry.ac.uk/ojs/index.php/joaw/article/view/165>.

Faulstich, Enilde (2000): « Principes formels et fonctionnels de la variation en terminologie. » *Terminology* 5.1: 93–106.

Firth, John Rupert (1957): *Papers in Linguistics* 1934–1951. London: Oxford University Press.

Francis, Gill (1986): *Anaphoric Nouns*. (Discourse Analysis Monographs 11). English Language Research, Department of English, University of Birmingham.

Francis, Gill (1994): "Labelling discourse: an aspect of nominal-group lexical cohesion." *Advances in written text analysis*. Ed. Malcolm Coulthard. London: Routledge. 83–101.

Firbas, Jan (1992): *Functional sentence perspective in written and spoken communication.* Cambridge: Cambridge University Press.

Giacomini, Laura (2015): "Context-dependent Variation of LSP Collocations: A Corpus-based Analysis." *Current Work in Corpus Linguistics: Working with Traditionally-conceived Corpora and Beyond.* Selected

Papers from the 7th International Conference on Corpus Linguistics (CILC 2015), 5–7 March 2015, University of Valladolid, Spain. Ed. Pedro A. Fuertes-Olivera et al. Procedia – Social and Behavioral Sciences 198. 40–148.

Gledhill, Christopher (1995): *Scientific Innovation and the Phraseology of Rhetoric. Posture, Reformulation and Collocation in Cancer Research Articles.* PhD Thesis, Aston University, UK.

Gledhill, Christopher (1997): « Les collocations et la construction du savoir scientifique. » *Anglais de spécialité* 15–18: 85–104.

Gledhill, Christopher (1999): "The phraseology of rhetoric, collocations and discourse in cancer research abstracts." *Knowledge and Discourse: Proceedings of the International Multidisciplinary Conference, 18–21 June 1996.* Eds. Colin Barron/Nigel Bruce. Hong Kong: University of Hong Kong.

Gledhill, Christopher (2011): "The 'Lexicogrammar' Approach to Analysing Phraseology and Collocation in ESP Texts." *Anglais de spécialité* 59: 5–23.

Halliday, Michael A. K. (1985): *Introduction to Functional Grammar*. London: Edward Arnold.

Halliday, Michael A. K. (1998a): "Language and knowledge: The 'unpacking' of text." *The Language of Science*. *Volume 5 in The Collected works of Michael A. K. Halliday.* Ed. Jonathan J. Webster. London/New York: Continuum (2004). 24–28.

Halliday, Michael A. K. (1998b): "Things and relations: regrammaticising experience as technical knowledge." *Reading Science: Critical and Functional Perspectives on Discourses of Science.* Eds. Jim Martin/Robert Veel. London: Routledge. 185–235.

Halliday, Michael A. K. (1995): "Language and Reshaping of Human Experience." *The Language of Science*. *Volume 5 in The Collected works of Michael A. K. Halliday.* Ed. Jonathan J. Webster. 2004. London/New York: Continuum. 7–23.

Halliday, Michael A. K. (1999): "The Grammatical Construction of Scientific Knowledge: The Framing of the English Clause." *The Language of Science*. *Volume 5 in The Collected works of Michael A. K. Halliday.* Ed. Jonathan J. Webster. London/New York: Continuum (2004).102–134.

Halliday, Michael A. K. (2003): "Written Language, Standard Language, Global Language." *World Englishes* 22.4: 405–418.

Halliday, Michael A. K./Hasan, Ruqaiya (1976): *Cohesion in English*. London: Longman.

Halliday, Michael A. K./Matthiessen, Christian M. I. M. (2014): *Introduction to Functional Grammar.* 4th edition (first published January 1st 1985). London: Hodder Arnold.

Harris, Zellig (1951): Methods in Structural Linguistics. Chicago: University of Chicago Press.

Harris, Zellig (1954): "*Distributional Structure.*" *Word 10.23: 146–162*.

Hoey, Michael (1991): *Patterns of Lexis in Text.* Oxford: Oxford University Press.

Humbley, John (2015): « La néologie dans les langues de spécialité. » *Neologia das Línguas Românicas*. Actes du IIe Congrès International de Néologie des Langues Romanes (CINEO II). 5–8 décembre 2011, Université de Sao Paulo, Sao Paulo, Brésil. Eds. Ieda Maria Alves/Eliane Simões Pereira. *HUMANITAS/CAPES*. 39–74.

Kübler, Natalie/Mestivier-Volanschi, Alexandra/Pecman, Mojca/Zimina, Maria (2016): « Exploitation quantitative de corpus de traductions annotés selon la typologie d'erreurs pour améliorer les méthodes d'enseignement de la traduction spécialisée. » *Actes des Journées internationales d'Analyse statistique des Données Textuelles (JADT 2016)* Volume 2. 7–10 juin 2016, Nice, France. UMR 7320 Bases, Corpus, Langage (CNRS – Université de Nice Sophia Antipolis). Eds. Damon Mayaffre/Céline Poudat/Laurent, Vanni,/Véronique Magri/Peter Follette. Presses de Fac Imprimeur. 731–741.

Kübler, Natalie/Pecman, Mojca (2012): "The ARTES bilingual LSP dictionary: from collocation to higher order phraseology." *Electronic lexicography*. Eds. Sylviane Granger /Magali Paquot. Oxford: Oxford University Press. 187–209.

Mathesius, Vilem (1928): "On linguistic characterology with illustrations from Modern English." *Actes du Premier Congres International de Linguistes, The Hague, 10–15 April 1928.* Leiden: Sijthoff. 56–63. (Reprinted in 1964, ed. Josef Vachek: *A Prague School Reader in Linguistics, Bloomington, IN: Indiana University Press.* 59–67.)

Mestivier-Volanschi, Alexandra (2015): "Productivity and Diachronic Evolution of Adjectival and Participial Compound Pre-modifiers in English for Specific Purposes." *Fachsprache* 37.1–2, 2–23.

Ormrod, Janet (2003): "Contrasts in modification and classification of nominal groups: native speakers versus French speakers writing in English." Communication presented at the *15th Euro-International Systemic Functional Linguistics Workshop*, July 20–23 2003. Leeds: University of Leeds.

Ormrod, Janet (2004): "Creation and subsequent usage of terms for discourse purposes in the scientific research article." *La production des textes specialises: structure et enseignement. Actes de Glat-Barcelona. 12–15 mai 2004.* Barcelona: Universitat Politècnica de Catalunya. 51–60.

Pecman, Mojca (2012): "Tentativeness in term formation: a study of neology as a rhetorical device in scientific papers." *Neology in Specialized Communication.* Eds. M. Teresa Cabré Castellví/Rosa Estopá Bagot/Maria C. Vargas Sierra. Special issue of *Terminology* 18.1: 27–58.

Pecman, Mojca (2014): "Variation as a cognitive device: how scientists construct knowledge through term formation." *Terminology* 20.1: 1–24.

Pecman, Mojca/Kübler, Natalie (2011): "ARTES: an online lexical database for research and teaching in specialized translation and communication." *Proceedings from International Workshop on Lexical Resources (WoLeR) 2011 at ESSLLI, 1–5 August 2011, Ljubljana, Slovenia*. ALPAGE/INRIA, Paris Diderot University, Paris: 86–93. (http://alpage.inria.fr/~sagot/pub/WoLeR_2011_proceedings.pdf)

Pic, Elsa/Furmaniak, Grégory/Hugou, Vincent (2013): « Étude comparative de quelques marqueurs de reformulation paraphrastique dans les articles de recherche et les articles de vulgarisation. » *Anglais de spécialité* 63: 75–92.

Sinclair, John McHardy (1980): "Some Implications of Discourse Analysis for ESP Methodology." *Applied linguistics* 1.3: 253–161.

Sinclair, John McHardy (1993): "Written discourse Structure." *Techniques of Description.* Eds. John McHardy Sinclair/Michael Hoey/Gwyneth Fox. London: Routledge. 6–31.

Sinclair, John McHardy/Mauranen, Anna (2006): *Linear Unit Grammar: Integrating Speech and Writing*. Amsterdam/Philadelphia: Benjamins.

Temmerman, Rita (2000): *Towards New Ways of Terminology Description: The Sociocognitive Approach.* Amsterdam/Philadelphia: Benjamins.

Tercedor, Maribel (2011): "The Cognitive Dynamics of Terminological Variation." *Terminology* 17.2: 181–97.

Williams, Geoffrey (1998): "Collocational networks: interlocking patterns of lexis in a corpus of plant biology research articles." *International Journal of Corpus Linguistic*s 3.1: 151–71.

Yamasaki, Nozomi (2008): "Collocations and colligations associated with discourse functions of unspecific anaphoric nouns." *International Journal of Corpus Linguistics* 13.1: 75–98.

Dr. Mojca Pecman
Centre de Linguistique Inter-langues, de Lexicologie,
de Linguistique Anglaise et de Corpus (CLILLAC-ARP) EA 3967
Université Paris Diderot
UFR EILA, case 7002
5 rue Thomas Mann
75205 Paris cedex 13
mpecman@eila.univ-paris-diderot.fr

Prof. Dr. Christopher Gledhill
Centre de Linguistique Inter-langues, de Lexicologie,
de Linguistique Anglaise et de Corpus (CLILLAC-ARP) EA 3967
Université Paris Diderot
UFR EILA, case 7002
5 rue Thomas Mann
75205 Paris cedex 13
cgl@eila.univ-paris-diderot.fr