English and French [NN]$_N$ lexical units:

A categorial, morphological and semantic comparison[1]

Pierre J.L. Arnaud and Vincent Renner

Abstract

This article presents a detailed classification of noun-noun nominal lexical units which shows that French and English have the same categories. It then contrasts French subordinative units with their English equivalents, and it appears that French units constitute less prototypical compounds, and their status as morphological or syntactic objects is a matter of debate.

> *Le français, en effet, possède divers genres de composition, mais craint, ce semble, d'en user.*[2]
>
> Arsène Darmesteter (1874: 241)

1 Introduction

There is a considerable amount of literature on compounding, and more loosely speaking, on the formation of COMPOSITE UNITS (a term we use to refer to noun-noun nominal ([NN]$_N$) units in general), with an increasing number of publications in recent years, for example the volumes edited by Lieber & Štekauer (2009) and Scalise & Vogel (2010). Various authors have proposed classification schemes (see Bisetto & Scalise 2005, Baroni *et al*. 2007 and Scalise & Bisetto 2009 for an overview), but typologies generally fail to include and characterize all productive semantic subtypes and to emphasize the similarities and dissimilarities between languages or language families. Using test sentences like those of Cruse (1986) and Riegel (1988), our goal in this article is to show that, beyond the oft-mentioned opposition between the strong compositive tendencies of Germanic languages and the much weaker ones

in the Romance languages, as well as that between the right-headedness of Germanic items and the left-headedness of Romance ones, the structure and semantics of [NN]$_N$ lexical units in English and French are parallel and the categorizations which apply to English are also applicable to French. We will, however, show that some categories exhibit marked differences between the two languages.

One main difference with much previous work lies in the fact that we do not refer in our classification to the concepts of endocentricity and exocentricity, which have been extensively used since Bloomfield (1933), but actually predate his treatise (see Noordegraaf 1989). In our view, the exocentricity of English and French [NN]$_N$ units is secondary to the word-formation process (Dressler 2006: 34 speaks of 'secondary head-hood' for exocentric compounds). Items like *ladder-back*,[3] *scissortail*, *micro-trottoir* 'mike pavement = a series of brief interviews of passers-by' or *gigot bitume* 'leg-of-lamb tar = a construction workers' picnic during which a carefully wrapped leg of lamb is cooked in a tank of hot tar' are more simply accounted for as resulting from metaphors or metonymies (see Benczes 2006: 4 and Bauer 2008a and 2008b for a similar view). Our current investigation is limited to French and English [NN]$_N$ items and until section 6 the term [NN]$_N$ UNIT is used without prejudging the theoretical status of the units in question as outputs of morphology. Our approach to classification rests on semantic headedness. Three different kinds of heads must be distinguished, however, if one is to account for exceptions to the 'normal' situation in English and French where the three heads are co-present in the same component, respectively N2 and N1, and we follow here the terminology of Scalise & Fábregas (2010). The semantic head of a unit is the component that corresponds to the superordinate concept, so when the lexical unit has a clear hyperonym, the head is normally the hyperonym (see Test T1 below). The semantic non-head specifies the head's meaning in an intersective way. The categorial head is the component whose word class is identical to that of the composite unit; for instance, *bluebell* being a noun, its categorial head is *bell*. Since we are dealing with [NN]$_N$ units, this notion is of minor interest here. The morphological head is the component whose morphological specifications (gender, declension class, etc.) correspond to those of the unit, and which receives number while governing agreement on other elements of the noun phrase and the sentence: the irregular plural of the morphological head *child* is found in that of *schoolchild*.

In English an exception to the transmission of the morphological specifications of the right-hand constituent to the unit is the plural form *goosefoots*. Examples of exceptions to semantic right-headedness are *endgame* and *peak oil*, whose semantics might lead to a left-headed interpretation, 'end of the game', 'peak of oil (production)'. French units with first components like *auto* or *radio* exhibit discrepancies in headedness. *Radio-crochet* 'radio hook = radio singing contest', *radiophare* 'radio lighthouse = electric beacon' and *auto-pont* 'auto bridge = fly-over' are semantically and morphologically right-headed. This is explainable by the fact that components like *auto* or *radio* are reanalyzed neo-classical elements that have become affixoids, that is 'affix-like morphemes [...] that still correspond to a lexeme' (Booij 2007: 86). *Autoradio* 'auto radio = car radio' and *bidonville* '(oil-)drum town = shanty town' are semantically right-headed, but their (masculine) gender is inherited from neither constituent. A possible explanation in the case of *bidonville* is that it results from extra-grammatical word-creation on the pattern of place-names like *Francheville*, built on the noun *ville*.F (*Francheville* is masculine, as shown by the toponym *Francheville le Haut* 'Francheville the High.M = Upper Francheville'). As to the gender of *autoradio*, it may be due to two facts: a) **autoradio*.F might be interpreted as 'a car for radio communication' with normal semantic left-headedness, and b) the French equivalent of *radio set* is *poste*.M *radio*, and *poste* is the hyperonym of *autoradio*. Only non-exceptional units will be discussed in the rest of the article.

The categories introduced below are prototype-based ones as, as we shall see, some semantically ambiguous units blur categorial limits. The units produced for the naming of concepts by the [NN]$_N$ pattern may arise only in specific instances of communication and so remain nonce formations; they may also be listed in the lexicon and therefore figure in dictionaries. Between these two extremes are units with few occurrences or belonging to specific terminologies. These sequences all result from the application of the same pattern and so will be used in this discussion as valid examples, contrary to non-attested sequences made up for metalinguistic purposes which are of dubious value in French, due to competition from other word-formation devices. In addition, French 'telegraphic style', that is the removal of closed-class words, results in NN sequences that occur on signs, in catalogues, job advertisements, etc., but are not lexical units. For this reason, only sequences found in full syntactic environments will be quoted as examples.

The organization of the paper is as follows. We introduce the binary top division between subordinative and coordinative [NN]$_N$ units in section 2, and then give a detailed description and analysis of the two categories in sections 3 and 4. We deal with more subclasses and some hard-to-categorize and / or deceptive units in section 5, compare the status of units in French with that in English in section 6, and summarize our main conclusions in section 7.

2 Subordinatives *vs.* coordinatives

Our top division rests on the presence or absence of a semantic nonhead-to-head relation. Units with such a relation, generally labelled as SUBORDINATIVE, denote a conceptual subcategory of the category denoted by the semantic head, which corresponds to N2 in English and N1 in French. This is evidenced by the test sentences in T1:

T1   (a)   An N1N2 is an N2.

         A space rocket is a rocket.

        (b)   *Un* N1N2 *est un* N1.

         Un pneu neige est un pneu.

         'A tyre snow is a tyre.'

The symmetrical test, T2, must return an unacceptable sentence:

T2   (a)   An N1N2 is an N1.

         *A space rocket is a space.

        (b)   *Un* N1N2 *est un* N2.

         **Un pneu neige est une neige.*

         'A tyre snow is a snow.'

Subordinative units are hierarchical in nature, and corresponding NNN units or the even more complex ones found in scientific and technical terminologies have internal structure due to the recursive application of the pattern (but see section 6 for restrictions about French):

(1)   (a)   [[railway]$_N$ station]$_N$

        (b)   [[birthday]$_N$ party]$_N$

(c)   [pearl [necklace]ₙ]ₙ

(d)   [[*crédit impôt*]ₙ *recherche*]ₙ

'credit tax research'

R&D tax credit

(e)   [*cigarette* [*bout filtre*]ₙ]ₙ

'cigarette tip filter'

filter-tip cigarette.

Units without a non-head-to-head relation are semantically coordinative, and generally known as COORDINATIVE UNITS, on account of the *and* that appears in their paraphrase and of their semantic similarity to *N and N* / *N et N* units, for example *rank and file* and *farces et attrapes* 'practical-jokes and traps = assorted tricks', as well as the lexicalized existence of two versions of the same denomination, as in *sale-leaseback* / *sale and leaseback* and *pieds-paquets* / *pieds et paquets* 'trotters (and) parcels = a dish of trotters and tripe'. They give equally acceptable or unacceptable sentences on T1 and T2:

T1   (a)   A hunter-gatherer is a hunter.

(b)   *Space-time is space.

(c)   *Un chasseur-cueilleur est un chasseur.*

'A hunter gatherer is a hunter.'

(d)   **Un service-volée est un service.*

'A serve volley is a serve.'

T2   (a)   A hunter-gatherer is a gatherer.

(b)   *Space-time is time.

(c)   *Un chasseur-cueilleur est un cueilleur.*

'A hunter gatherer is a hunter.'

(d)   **Un service-volée est une volée.*

'A serve volley is a volley.'

In contrast to the situation of NNN subordinatives, coordinative NNN units have a semantically flat, non-hierarchical structure, which suggests that they have a radically different nature:

(2) (a) feller buncher skidder

(b) writer-director-producer

(c) bridge-tunnel-bridge

(d) *moissonneuse-batteuse-lieuse*

'harvester thresher binder'

(e) *auteur-compositeur-interprète*

'writer composer singer'

(f) *riz-pain-sel*

'rice bread salt'

quartermaster.

The preceding examples are lexicalized; triple and longer coordinative sequences, however, are mostly nonce formations:

(3) (a) But there is something disturbing about the collective rejection-embrace-elevation of Susan Boyle.

(b) L'*avocat-écrivain-journaliste-essayiste* Emmanuel Pierrat.

'lawyer writer journalist essayist'

Lastly, mixed types associating coordinative and subordinative relations are possible, as in *fairy godmother*, *peanut butter jelly sandwich*, *maître-nageur sauveteur* 'master swimmer lifeguard' and *policier homme-grenouille* 'policeman man-frog = police diver'.

A number of authors, for example Fabb (1998), Olsen (2004) and Dressler (2006), have put forth a dual classification of [NN]$_N$ units. Scalise & Bisetto (2009), by contrast, present a threefold classification. Applying Lieber's (2004, 2009) model of lexical representations, their three higher-level classes (SUB, ATAP and COORD) correspond to SUBORDINATION, ATTRIBUTION-APPOSITION and COORDINATION, which they justify by the fact that these are the three categories of relations existing in syntax. Like Scalise & Bisetto's classification, a dual division based on the presence / absence of a semantic nonhead-to-head relation can be extended to other word classes, as in Table 1 (note that subordinative verbs are missing in French outside neoclassical compounding of the *pyrograver* (pyro- + graver$_V$) type ['pyro- + engrave = to

wood-burn']). It also has the advantage of taking into account important morphological and phonological facts found in a number of languages: in particular, contrary to subordinatives, coordinatives in, for example, English and German predominantly have phrasal, that is, right stress;[4] German coordinatives normally do not include linking elements (Becker 1992, Neef 2009); also, the sandhi phenomena attested in the compounding process for instance in Japanese (*rendaku*), Korean (*sai-sios*) and Malayalam are blocked on semantic grounds in the case of coordinatives (Kim 2001).

	nominal	adjectival	verbal
subordinative	*piggy bank*	*squeaky-clean*	*shrink-wrap*
	appui feu 'support fire' fire support	*ivre-mort* 'drunk dead' dead drunk	
coordinative	*bridge-tunnel*	*manic-depressive*	*freeze-dry*
	guide-conférencier guide lecturer	*obsessif-compulsif* obsessive compulsive	*copier-coller* copy paste

Table 1. A dual semantic classification of composite lexical units

3 Subordinatives

In this section we examine the subdivisions of the subordinative class. Two subclasses can be distinguished by observing their behaviour relative to T3, a test of analogy (the French examples are exact equivalents of the English ones):

T3 (a) An N1N2 is an N2 which is like an N1.

      A crocodile clip is a clip which is like a crocodile.

      *A shower gel is a gel which is like a shower.

  (b) *Un* N1N2 *est un* N1 *qui est comme un* N2.

      Une pince crocodile est une pince qui est comme un crocodile.

      **Un gel douche est un gel qui est comme une douche.*

Those units that return acceptable utterances on T3 all have affinities with metaphor in that only a subset of the perceptual, behavioural or functional features in the representation of the non-head serves to isolate a subcategory of the head. In the case of *crocodile clip / pince crocodile*, only the snout and the teeth of the crocodile provide a description for the device in question, while its scaly skin, aquatic life and predatoriness are overlooked in the compositive process. As the semantic relation of modification in this subclass consists in the attribution of features of the non-head to the head, the label ATTRIBUTIVE is adequate. This attribution results in semantic intersection between the representations of N1 and N2. Other examples are *beltway*, *bullet train*, *hermit crab*, *kidney bean*, *chapeau melon* 'hat melon = bowler hat', *crabe yéti* 'crab yeti = *Kiwa hirsuta*', *pommes allumettes* 'potatoes matchsticks = thin-cut fries', *ville champignon* 'town mushroom = boomtown' and *voiture pie* 'car magpie = panda car'.

The other subordinatives, like *shower gel* and *gel douche* above, do not exhibit analogy in their non-head-to-head relation, whose nature is much more difficult to describe in simple terms, so the term RELATIONAL can be retained for them. On the surface, the modification relation provides the semantic intersection that subcategorizes the head as in the case of attributives. A difference with attributives is that the meaning can be represented by means of a predication, with two cases. In the more frequent case, both N1 and N2 are arguments or complements of the predication:

(4)   seaweed

        location$_{(SPACE)}$ (weed, sea)

(5)   fire alarm

        be-against (alarm, fire)

(6)   nightshift

        location$_{(TIME)}$ (shift, night)

(7)   *bébé éprouvette* 'baby test-tube = test-tube baby'

        come-from (baby, test-tube)

(8)   *menu enfant* 'menu child = child menu'

        be-for (menu, child)

(9)   *camion benne* 'truck bin = dump truck'

have (truck, bin).

In the second, less frequent case, one of the components, a base verb form or deverbal noun in English or a deverbal noun in French, corresponds to the predicator, the other component being the external or internal argument:

(10)   landslide

         slide (land)

(11)   landfill

         fill (x, land)

(12)   *équipe action* 'team action = action team'

         act (team)

(13)   *suivi client* 'follow-up client = client follow-up'

         follow up (x, client).

Relational units range from items with a relatively high degree of transparency, for example *forest fire*, *shoe polish*, *table top*, *bas nylon* 'stocking nylon = nylon stocking', *ingénieur béton* 'engineer concrete = concrete engineer' and *timbre poste* 'stamp post = postage stamp', to situation-dependent novel units that can only be interpreted pragmatically, like Downing's (1977) oft-quoted example 'apple-juice seat', as well as lexicalized units semantically underpinned by a complex scenario, such as *banana republic* 'a small country in the tropics which is economically dependent on a single export commodity and run kleptocratically', *bingo wings* 'flabs under the upper arms of elderly women', *soccer mom* 'a middle-class suburban stay-at-home woman with children', *appartement ravioli* 'flat ravioli = a flat in which illegal migrants prepare unsanitary food for Chinese restaurants' and *argent braguette* 'money trouser-fly = child allowance money' (in overseas regional French), *retraite cocotier* 'retirement-pension coconut-tree = the high pension of a metropolitan civil servant who retires in the Pacific Islands'. But even in these latter cases, the relation in itself is rather uncomplicated when the intervening metonymies are removed:

(14)   banana republic

         produce (republic, banana)

(15)   *retraite cocotier*

location$_{(SPACE)}$ (retirement-pension, coconut-tree).

Conversely, as has often been noted, even fairly obvious units may appear ambiguous when one goes beyond their lexicalized sense: *forest fire* might, after all, denote a fire made in a forest for signalling or cooking purposes. Another notable characteristic of the class is that several relations may co-exist, a case of 'promiscuity' in the terminology of Jackendoff (2009). This problem was noted by Mätzner (1860: 472) and Darmesteter (1877: 160) was wondering whether French *café-concert* denoted a café which was a concert hall or else a café in which concerts took place, and, for example, names of buildings or rooms like *interrogation room* or *local poubelles* 'room dustbins = bin room' often combine the location relation with the telic one which is common to names of artefacts. Selkirk (1982: 25) summarized the problems involved in categorizing the relations:

> […] the range of possible semantic relations between the head and non-head is so broad and ill defined as to defy any attempt to characterize all or even a majority of the cases.

It is not possible to survey here all the classifications that have been put forth, but they range from very restricted, abstract lists like those of Hatcher (1960) for English and Noailly (1990) for French, with four similar categories (modulo head positions), namely N1 origin of N2, N1 goal of N2, N1 in N2, N2 in N1, to much finer-grained ones like that of Arnaud (2003: 65-7), who isolated 54 categories in French relational [NN]$_N$ units. Between such extremes are, for instance, Lees' (1970) list of 13 'generalized verbs' complemented with 6 non-verbal relations, the 9 'recoverable predicates' of Levi (1978), or Jackendoff's (2009) 13 'basic functions'.

From an onomasiological perspective, attribution in French can only be found in [NN]$_N$ units while the relational modification is also found in prepositional units like *voiture de course* 'car of race = racecar' or *patin à roulettes* 'skate at / to rollers = roller skate', or in noun + relational adjective sequences like *conseil municipal* 'council municipal = city council', which are much more frequent than NN items. A prepositional unit like *scie à ruban* 'saw at / to ribbon = bandsaw' is not attributive as the preposition expresses the merological relation (N2 is a part of N1); analogy is present in the metaphor on N2 (the blade is like a ribbon). Similar examples are *piano à queue* 'piano at / to tail = grand piano', *verre à pied* 'glass at / to foot = stem glass' and *serpent à sonnette* 'snake at / to bell = rattlesnake'.

Finally, English and French do not have watertight noun and adjective classes and they exhibit NN sequences[5] which superficially might be taken for subordinative composites, but in which the non-head functions like a syntactically attributive adjective in several respects. Examples of such non-heads are:

(16) (a) freak, ghost, key, king, legend, mammoth, marathon, miracle, orphan, pocket, token

      (b) *choc* 'shock', *clé* 'key', *culte* 'cult', *éclair* 'lightning', *fantôme* 'ghost', *miracle* 'miracle', *miroir* 'mirror', *modèle* 'model', *mystère* 'mystery', *phare* 'lighthouse', *pilote* 'pilot'.

Gushchina-Magno (2009: 74-86) reviews the problems involved in defining the status of such French N2s, and it appears that many of the possible criteria, like adjectival derivation or the impossibility of using N2 in a predicative noun phrase, as in (17) and (18), are applicable only to a subset of the units.

(17) *culte*   >   *cultissime*

      'cult'   'very-cult'

(18) *une robe olive*   >   **cette robe est une olive*

      'a dress olive [= olive dress]'   'this dress is an olive'

A general semantic difference with composite lexical units is that the modification does not have a categorizing effect but a qualifying or describing one; the resulting NN sequences are not 'names' (Riegel 1991) and do not denote stable concepts corresponding to categories of entities. This results from a reduction of the semantic representation of the non-head, from which one feature is promoted, as in the non-prototypical French adjective *marron* 'chestnut = brown' where only the feature COLOUR remains from the representation of the noun it evolved from:

(19) Is blanket firing of teachers legal in states other than Rhode Island?

(20) *Toyota a enregistré des bénéfices record au dernier trimestre de 2006.*

      'Toyota made record profits in the last quarter of 2006.'

Most of these NN sequences include an attributive relation, but some, like *cult* / *culte* or *pocket* are relational in nature. More importantly, some of these nouns exhibit a degree of adjective-like syntactic behaviour, such as predicative occurrences, modification by an intensity adverb, etc., which varies from noun to noun (see Arnaud 2010 for English), and this behaviour is impossible in the case of composites:

(21) holding indefinitely profiles from everyone who was arrested was 'blanket and indiscriminate'

(22)   Councils in Northern Ireland have reported <u>almost blanket</u> compliance since the new smokefree

law came in April.

(23)   *une affluence, <u>quasi record</u> pour un samedi matin,* […]

'a crowd, quasi record for a Saturday morning'

almost a record crowd for a Saturday morning

(24)   *L'année 2009 <u>s'annonce record</u>.*

'The year 2009 itself-announces record'

The year 2009 should be a record year.

Contrast this with the equivalent attributive lexical units *catfish / poisson-chat*:

(25)   *This is a fish that has remained cat in spite of its untypical habitat.

(26)   **C'est un poisson qui est resté chat malgré son habitat atypique.*

In some cases, however, no semantic reduction is detectable, nor is the modifier attested in non-attributive adjectival positions. The multiplication of NN sequences with identical N2s might give the impression that these N2s function like adjectives, but there is no objective reason not to consider the corresponding items as composite lexical units, as in the following examples:

(27)   (a)   radar albedo, radar antenna, radar ceiling, radar clutter, radar detection, radar discretion, radar echo, radar horizon, radar noise, radar signal, radar signature

         (b)   *baril papier* 'barrel paper = oil shares', *bulletin papier* 'ballot paper = paper ballot', *convocation papier* 'notice paper = paper notice', *copie papier* 'print-off paper = hard copy', *exemplaire papier* 'copy paper = paper copy', *journal papier* 'newspaper paper = printed newspaper', *pierre papier* 'stone paper = real-estate shares'.

However, an apparently intractable problem with NN sequences including adjectivized nouns is that these sequences pass Test T3 when the qualification is semantically attributive. When the other possible criteria do not apply, one has to fall back on the two above-mentioned, somewhat subjective notions of the semantic reduction of the non-head and qualification *vs.* subcategorization to distinguish lexical units from occasional noun phrases. This blurs category limits somewhat, and makes a prototypical approach desirable.

4 Coordinatives

Establishing a watertight distinction between subordinative and coordinative units is a thorny task. The fact that the semantic relation holding between the constituents can be paraphrased by AND in English / ET in French is a necessary but non-sufficient condition for membership in the coordinative category. In our view, membership also entails that the constituents are co-hyponyms, which leads us to discard items such as *manservant*, *child bride*, *femme médecin* 'woman doctor' or *bébé phoque* 'baby seal' from the category (see section 5 below), and that the elements are on an equal grammatical footing. Linear order is conditioned by various factors, pragmatic and phonological: the constituent referring to an event or action first in time will be ordered first (Olsen 2001), as in *dinner dance*, *washer-dryer*, *juillet-août* 'July August' or *service-volée* 'serve and volley'; the second constituent tends to be syllabically longer than the first one. Non-hierarchy is reflected in the possibility of linear reversibility for the constituents, as shown in T4:

T4   (a)   N1N2 could be named N2N1.

            Oxidation-reduction could be named *reduction-oxidation*.

            *A mushroom town could be named a town mushroom.

    (b)   N1N2 *pourrait être dénommé* N2N1.

            La chasse-cueillette pourrait être dénommée cueillette-chasse.

            'Hunting gathering could be named gathering hunting.'

            La gomme-résine pourrait être dénommée résine-gomme.

            'Gum resin could be named resin gum.'

            *Un oiseau mouche pourrait être dénommé mouche-oiseau.

            'A bird fly [= hummingbird] could be named a fly bird.'

The property of linear reversibility sometimes leads to the lexicalization of the two units as synonyms (for example *lease-lend* / *lend-lease*, *amiante-ciment* / *ciment-amiante* 'asbestos cement' or *casque-micro* / *micro-casque* 'headphone mike = headset'), which can be seen as a sufficient condition for coordinative status. An exception to the validity of T4 holds for asynchronous items (for example *murder-suicide*, *hunter killer*, *cadrage-débordement* 'committing swerving = side step' [in rugby] or *juillet-août*), whose

constituents refer to sequential events, actions or abstractions, but asynchrony itself may be taken as evidence of non-subordinative status.

The divisions within the category are somewhat uncertain, but they are less crucial than in the case of subordinatives, which may partially explain why the categorization of coordinatives is usually not as detailed as that of subordinatives in the morphological literature. Because of the complexity of the problem, a multiple-prototype approach is useful. Three prototypes can be identified. They correspond to types of coordinatives that were, to the best of our knowledge, first brought to light by Hatcher (1951: 32): 'one entity seen as two', 'two entities seen as one' and a 'mixture' of two entities.

We term the first type MULTIFUNCTIONAL UNITS. They are felicitously inserted in T5:

T5   (a)   An N1N2 is an N1 who / which is also an N2.

            A fighter-bomber is a fighter which is also a bomber.

            A student-athlete is a student who is also an athlete.

     (b)   *Un* N1N2 *est un* N1 *qui est aussi un* N2.

            Un boucher-charcutier est un boucher qui est aussi un charcutier.

            'A butcher pork-butcher is a butcher who is also a pork-butcher.'

            Une librairie-papeterie est une librairie qui est aussi une papeterie.

            'A bookshop stationer's is a bookshop which is also a stationer's.'

A multifunctional unit is composed of constituents which each designate a function of the entity in question. This type of unit frequently denotes occupations (for example *broker-dealer*, *owner-occupier*, *player-manager*, *speaker-hearer*, *animateur-producteur* 'TV presenter / producer', *député-maire* 'MP mayor', *ouvrier-paysan* 'worker peasant' or *layetier-emballeur* 'trunk-maker packer') and artefacts (for example *transmitter-receiver*, *sofa bed*, *porte-fenêtre* 'door window = French window' or *stade-vélodrome* 'stadium cycle-stadium').

The second type is hybrid units. These are felicitously inserted in T6:

T6   (a)   N1N2 is a blend / hybrid of N1 and N2.

            Blues-rock is a blend of blues and rock.

            The troutperch is a hybrid of the trout and the perch.[6]

(b)　N1N2 *est un mélange / hybride de* N1 *et de* N2.

　　La gomme-résine est un mélange de gomme et de résine.

　　'Gum resin is a blend of gum and resin.'

　　Un roman-poème est un hybride de roman et de poème.

　　'A novel poem is a hybrid of a novel and a poem.'

They denote mainly abstractions (for example *rap metal*, *southwest*, *roman-poème* and *nord-est* 'northeast') and substances (for example *gum resin*, *tarmacadam*, *toxin-antitoxin*, *gin-tonic* 'gin and tonic', *whisky-soda* 'Scotch and soda' and *eau-méthanol* 'water methanol').

The third type is additional units, which are felicitously inserted in T7:

T7　　(a)　An N1N2 is an N1 plus an N2.

　　A tractor-trailer is a tractor plus a trailer.[7]

　　Space-time is space plus time.

　　(b)　*Un* N1N2, *c'est un* N1 *plus un* N2.

　　Une saisie-vente, c'est une saisie plus une vente.

　　'A seizure sale is a seizure plus a sale.'

　　Le canoë-kayak, c'est le canoë plus le kayak.

　　'Canoeing kayaking is canoeing plus kayaking.'

Unlike multifunctionals and hybrids, additionals designate entities formed by the juxtaposition, and not the fusion, of the *denotata* of the constituents. They denote mainly artefacts (for example *fridge-freezer*, *tractor-trailer*, *châssis-carrosserie* 'chassis body' and *micro-casque*) and events (for example *dinner dance*, *rape-murder*, *camping-caravaning* 'camping and trailering', *chasse-cueillette* 'hunting and gathering', *douche-massage* 'shower massage' and *recherche-développement* 'research and development').

Not all coordinatives can be neatly pigeonholed into one of the three categories. *Hammer axe / marteau piolet*, for instance, is functionally multifunctional, but perceptually hybrid; *point-virgule* 'full-point comma = semi-colon' is perceptually additional, but semantically hybrid. In some cases, two interpretations may co-exist: *washer-dryer* and *collège-lycée* 'middle-school high-school' will be

classified as additionals in case of juxtaposition of the *denotata* and as multifunctionals in case of fusion of the *denotata*.

5 More subclasses

In a classificatory framework different from ours, that of Bisetto & Scalise (2005), Arcodia *et al.* (2009) note that, from a purely semantic point of view, the distinction between attributive-appositive and coordinative items is not clear. Similarly, in our classification as it now stands, units like *guide dog* and *loi-cadre* 'law framework = framework law' are problematic: they give acceptable sentences on both T1 and T2 and so might be confused with coordinatives:

T1 A guide dog is a dog.

T2 A guide dog is a guide.

T1 *Une loi cadre est une loi*.

T2 *Une loi cadre est un cadre*.

It can, however, be demonstrated that they are different from coordinatives. First, N1 and N2 are not co-hyponyms. Then, the generic in dictionary definitions is N2 in English and N1 in French:

(28) A dog that has been trained to lead a blind person. (Soanes & Stevenson 2011)

(29) *Loi dont les dispositions générales doivent servir de cadre à des textes d'application*. (Rey-Debove & Rey 2011)

   'A law whose general provisions must be used as a framework for the implementing provisions.'

Hyperonymic anaphora repeats respectively N2 and N1, as in the following test sentences:

(30) Guide dogs are trained for months by sighted volunteers. Only when they are ready for service are the dogs / *the guides presented to those who need them.

(31) *Une loi-cadre sur les universités a bien été votée en 1995, mais cette loi / *ce cadre n'a jamais été appliqué(e)*.

   'A framework law on universities was voted in 1995, but this law / *this framework was never applied.'

We therefore have unequivocal heads: the units are subordinatives, which is corroborated by the left-hand stress of the majority of such English units. Why is it, then, that T2 returns acceptable sentences? The copula in them shows that we have a case of attributive modification, but T3 gives slightly infelicitous sentences:

T3   (a)   [?]A guide dog is a dog which is like a guide ('A guide dog is a dog which is a guide.' sounds better).

        (b)   [?]*Une loi-cadre est une loi qui est comme un cadre* ('*Une loi-cadre est une loi qui est un cadre.*' sounds better).

This shows that the modification is not unequivocally analogical, contrary to the examples of attributive compounding in section 3. There seem to be two reasons for this: the modifier may be a catachrese, which appears as a separate sense in dictionaries independently of the compositive process, so only a subset of the original features subsists in its representation (this is the case of *guide* in *guide dog* and *cadre* in *loi-cadre*), or else its meaning may be general enough, like that of *coil* in *coil spring* and *poids* in *barrage-poids* 'dam weight = gravity dam'. This means that the attributive subordinative category needs to be refined: beside analogical units, we need another subclass, for which we suggest the term EQUATIVE on account of the copula present in paraphrases. As the difference between analogicals and equatives rests on metaphor-like attribution *vs.* literal attribution, one can expect there to be intermediate cases, depending on the perception of the analogy. Other examples of equative attributives are *fuel oil*, *junk food*, *tree fern*, *dalle-plancher* 'slab floor = floor slab', *légume-racine* 'vegetable root = root vegetable' and *récif-barrière* 'reef barrier = barrier reef'.

At first sight *oaktree*, *codfish*, *pierre ponce* 'stone pumice = pumice stone' and *thon germon* 'tuna albacore = *Thunnus alalunga*' are equatives as they give the same results on T1 and T2. The results on T3, however, are very clearly unacceptable, whereas those of equatives are only slightly infelicitous.

T3   (a)   *An oaktree is a tree which is like an oak.

        (b)   **La pierre ponce est une pierre qui est comme la ponce.*

The modification is attributive in nature, but in a special way; instead of being a noun with independent denotation, whose representation modifies that of the head to create a subcategory, the modifier is a hyponym of the head, denoting the same entity as the whole unit:

(32)   oak   =   oaktree

(33)   *ponce*   =   *pierre ponce*.

The composite unit is more explicit than the modifier used separately as it includes the hyperonym, and this may explain the existence of these units, which alternate in discourse with the simplex form. With regard to their analogy with the names of biological taxonomy, we suggest the term GENERIC-SPECIFIC for this subclass of attributives.[8]

Items like *sawfish*, *lyrebird*, *poisson-scie* 'fish saw = sawfish' and *oiseau-lyre* 'bird lyre = lyrebird' are also deceptive. On superficial examination, they might appear as analogical attributives. Test T8, however, which consists in forming a noun phrase with the modifier postmodified by the whole unit, is revealing:

T8   (a)   The N1 of an N1N2 […]

              The saw of a sawfish [is a rostrum it uses to detect prey].

       (b)   *Le* N2 *d'un* N1N2 […]

              La scie d'un poisson-scie [*est un rostre qui lui sert à détecter ses proies*].

Compare with true analogicals:

(34)   catfish: *The cat of a catfish […]

(35)   *poisson-chat*:   **Le chat d'un poisson-chat* […].

This is due to the fact that T8 tests the OF abstract relation (or, conversely, the HAVE relation), which dominates the merological relation, one of the many modification relations found in relational units (see supra, section 3). *Sawfish* and *poisson-scie* are not analogicals, but relationals with the merological relation and a metaphorized non-head. They tend to be interpreted as analogicals because of the salience of the metaphor. We need not modify our classification on account of such items, which exist in very small numbers anyway and are probably not perceived as a category by speakers, but they do introduce a limited intersection between relationals and analogicals.

Lastly, we have a set of difficult customers in which N1 and N2 are not co-hyponyms and which at first sight seem to be equatives, but also strongly resemble coordinatives: *child bride*, *child soldier*, *detective inspector*, *woman doctor*, *apprenti-cuisinier* 'apprentice cook', *enfant soldat* 'child soldier' and *femme médecin*. One of the nouns frequently denotes a status or age group with respect to the category denoted by the other noun. That their headedness is unclear appears in the fact that *child soldier* and *enfant-soldat*, *woman doctor* and *femme médecin*, with identical *denotata*, have their constituents in the same order in spite of the different positions of the heads in the two languages. Most English units have right-hand stress, which would seem to locate them within the coordinative class. Determining objectively where their semantic heads actually are is difficult, which is reflected in the discordant dictionary definitions given for a unit like *child prodigy*. Several lexicographic sources, for example Macmillan, Cambridge and Longman, define *child prodigy* as denoting primarily a child, ('a child who […]'), but one, WordNet, has a definition starting with the words 'a prodigy […]'.[9] Tests with anaphora return mixed results. Our informants unanimously prefer *child* to *soldier* in hyperonymic anaphors to *child soldier*, they reject passages in which *child bride* is taken up by either *child* or *bride*, or else state that they have a slight preference for *bride*, and a clear majority prefer *inspector* rather than *detective* to refer back to *detective inspector*. Informants prefer *enfant* in anaphors to *enfant-soldat* and *apprenti* in the case of *apprenti-cuisinier*, which points to their possible status as equative attributive subordinatives, but they reject anaphors to *femme-médecin* by either *femme* or *médecin*. These units evidently do not exhibit unitary behaviour and, in a prototypical perspective, it seems wiser not to set up a separate class for them and, like Arcodia *et al*. (2009), who place them on a continuum, to consider that they occupy different positions on the cline between equatives and multifunctional coordinatives.

The main types and subtypes of [NN]$_N$ lexical units having been dealt with, we can now present a diagram of our classification in Figure 1. Dotted double arrows are used to indicate the existence of a continuum between categories, and the dotted line is used to indicate that the subdivisions within the category of coordinatives are somewhat uncertain.[10]
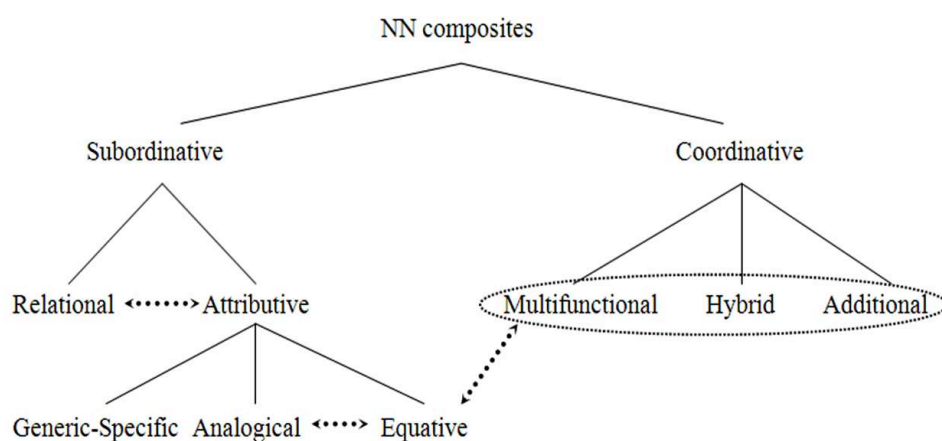
Figure 1. Categories of NN composite lexical units

As said in section 1, we consider that tropes can be disregarded in the classification of [NN]$_N$ lexical units as they are external to the word-formation process. It is, however, important to determine which categories can be the object of a trope. Metaphorical and metonymic units seem to be restricted to the following three subclasses (the literal meaning may either co-exist with the figurative one or have been displaced by it):

– relational subordinatives (for example *bookworm*, *bullhorn*, *paperback*, *sand dollar*, *tennis elbow*, *année-lumière* 'year light = light year', *cheval vapeur* 'horse steam = horsepower', *bœuf-carottes* 'beef carrots = a nickname for officers of the French police internal affairs division', *emploi jeune* 'job young-adult = a person holding a goverment-funded job reserved for young adults' and *poids plume* 'weight feather = featherweight boxer; also lightweight'), sometimes with the merological relation (for example *bullhead*, *turtleneck*);

– analogical attributive subordinatives (for example *egghead*, *pintail*, *homme orchestre* 'man orchestra = lit. one-man band; fig. a person having many diverse activities', *poisson pilote* 'fish pilot = lit. pilotfish; fig. a person who acts ahead and covertly for someone else' and *voiture-balai* 'car broom = lit. broom wagon; fig. a person or organization in charge of rescuing people or organizations in trouble');

– additional coordinatives (for example *broomball*, *figure-ground*, *penny-farthing*, *aldéhyde alcool* 'aldehyde alcool = a compound containing aldehyde and alcohol functional groups',

bains-douches 'baths showers = public baths', *Paris-beurre* 'Paris butter = a sandwich with butter and Paris ham' and *railroute* 'rail road = combined rail / road transport').

6 Beyond classification

We have shown in the preceding sections that, from a classificatory perspective, English and French [NN]$_N$ units are similar. In this section we compare them with respect to their status. From a lexicalist perspective, words have lexical integrity and, as Anderson (1992: 84) wrote, '[t]he syntax neither manipulates nor has access to the internal structure of words'. English compounds are known to be problematic in this respect, being the least prototypical outputs of morphology. We now investigate the situation of French units.

An important criterion that could help us answer the question of their morphological wordhood is unfortunately unavailable for comparison purposes: while stress is not an entirely reliable proof of compoundhood in English, there is a tendency for subordinatives to be stressed on N1, so when an NN unit is left-stressed, we can be reasonably sure that it is a compound, that is a morphological word, as syntactic phrases bear right stress (the reverse is not true, however).[11] French, by contrast, does not have lexical stress.

Being lexemes, compounds constitute bases for derivation, as in *penmanship* (*penman* + *-ship*) or *watercolourist* (*watercolour* + *-ist*). The suffix in *motocrottiste* 'motorcycle faeces -ist = the operator of a motorcycle equipped to remove dog faeces from pavements' can attach to phrases as well as words (for example *je m'en foutiste* 'I don't care -ist = someone with a couldn't-care-less attitude', *fil-de-fériste* 'thread of iron -ist = tightrope walker' and *droits-de-l'hommiste* 'rights of man -ist = human-rightist'), so this unit cannot be retained as evidence of the morphological status of its base, but *capital-risqueur* 'capital risk -er = venture capitalist' or *mot-valisage* 'word portmanteau -age = lexical blending' are more useful examples. They are, however, isolated examples in our data, as derivatives from French [NN]$_N$ units are hard to find. At first sight, pairs like (36) to (39) would seem to prove the possibility of derivation from the head, so of internal derivation, and thus cast a doubt on the status of French [NN]$_N$ units as morphological outputs. All the derived N1s are however independently attested, so we cannot

exclude that they were used directly as bases in the formation of the composite units, by 'analogical extension' (Haspelmath & Sims 2010: 127) from the first unit in each pair, as illustrated in (40), in which case these items tell us nothing about the status of French [NN]$_N$ units.

(36)   *chêne liège ~ chênaie liège*

         'oak cork ~ oak-forest cork'

         cork oak ~ cork oak forest

(37)   *chèque service ~ chéquier service*

         'cheque service ~ cheque-book service'

         service voucher ~ service voucher book

(38)   *contrôle qualité ~ contrôleur qualité*

         'control quality ~ controller quality'

         quality control ~ quality controller

(39)   *épargne retraite ~ épargnant retraite*

         'savings retirement ~ saver retirement'

         retirement savings ~ retirement saver

(40)   [*chèque*]$_N$ + [*-ier*] ⇒ [*chéquier*]$_N$
                        ↙   ↘
         [*chèque*]$_N$ + [*service*]$_N$ [*chéquier*]$_N$ + [*service*]$_N$
                  ⇓   ⇓
            [*chèque service*]$_N$ [*chéquier service*]$_N$

One aspect of syntactic impenetrability is anaphoric islandhood; pronominal anaphoras cannot have the nonhead of a compound as antecedents, as is apparent in the following test sentences:

(41)   *The sea$_i$lions are disappearing. It$_i$ is too heavily polluted.

(42)   *Il y a pénurie de timbres-poste$_i$. En fait, elle$_i$ n'en a pas fait imprimer assez.*

         'There is a dearth of stamps-post$_i$ [= postage stamps]. In fact, it$_i$ [= the post-office] hasn't had enough printed.'

Ward *et al*. (1991), however, present real-life examples of anaphora to the non-head in English like the following:

(43)   In the distance, we heard the sound of an ambulance$_i$ siren. Within a minute or so it$_i$ arrived and

        stretcher bearers took the boy away.

Also, the following French test utterances may not be absolutely felicitous, but even presented in isolation

they are accepted by most informants:[12]

(44)   *L'impôt sécheresse$_i$ va être reconduit pour un an. Il est vrai qu'elle$_i$ a été encore pire cette*

        année.

        'The tax drought$_i$ [= drought tax] will be extended for a year. It is true that it$_i$ has been even

        worse this year.'

(45)   *Le contrôle qualité$_i$ a été négligé. Désormais, on ne peut pas être sûr qu'elle$_i$ reste constante.*

        'Control quality$_i$ [= quality control] has been neglected. From now on, one cannot be sure that it$_i$

        remains constant.'

According to Ward *et al*. (1991), such examples challenge the dogma of anaphoric islandhood and its

syntactic explanations. These authors suggest that the correct explanation is pragmatic and discoursal:

anaphora is possible if the non-head antecedent is accessible in the hearer's discourse representation.

Observing similar data from Italian, Gaeta & Ricca (2009) conclude that syntactic islandhood is on a

gradient and correlates more with degree of lexicalization than with morphological status. At any rate,

French units do not seem to behave differently from English ones in this respect, so we need to look

further for clear differences.

English compounds have often been considered problematic with respect to the morphology-syntax

interface as they can include supposedly syntactic outputs in so-called phrasal compounds like *wine and*

cheese party, *fire and forget weapon*, *line-of-battle ship*, *bird-of-paradise feathers* and *postage due*

stamp. French not only freely allows similar constructions (for example *conseiller presse et*

communication 'adviser press and communication = press and communication adviser', *responsable*

hygiène et sécurité 'officer health and safety = health and safety officer', *assurance perte de loyers*

'insurance loss of rent = rent loss insurance', *plan épargne salariale* 'plan savings employee = employee

savings plan', *tube modes doux* 'tube means mild = a tunnel accessible to environment-friendly transport',

plan grand froid 'scheme great cold = a scheme to provide homeless people with shelter' and *congé*

longue maladie 'leave long illness = long-term sick leave'), but its [NN]$_N$ units manifest other influences from syntax, in particular on pluralization.

In both languages, the pluralization of composite units is marked on the head,[13] so French units have an internal plural while that of English ones may be analyzed as external, as shown by the brackets in the following examples:

(46)   attributives: *animaux*°[14] *machines vs.* [*bellflower*]s

                                   'animals machines'

                                   machine animals

(47)   relationals: (a)   *chevaux° vapeur vs.* [*mailbox*]es

                                   'horses steam'

                                   horsepower

            (b)   *œufs° cocotte*

                                   'eggs pot'

                                   shirred eggs

            (c)   *ponts-canal°*

                                   'bridges canal'

                                   navigable aqueducts[15]

(48)   coordinatives: *députés-maires vs.* [*owner-occupier*]s

                                   'MPs mayors'

Internal plurals, however, do not constitute a decisive argument against compoundhood, as argued by Booij (2010: 174) on the grounds that the 'word-internal operation of morphological rules is not excluded by the Lexical Integrity Principle'.[16] These examples show another contrast: all English units have a single plural mark; in French there is a tendency for relationals to have a single plural mark while attributives and coordinatives have two (for example *bateaux boeufs°* 'boats oxen = a type of trawler', *tuiles canaux°* 'tiles canals = Roman tiles', *romans journaux°* 'novels diaries = diary novels' and *chiens-loups* 'dogs wolves = (informal) German shepherds', *plombiers-zingueurs* 'plumbers roofers', *batteries-fanfares* 'drum-corps fanfares').[17] A third difference concerning plurals is that the non-head of an English

subordinative unit is generally not pluralized, even if it denotes several entities: a passenger train is for passengers, a bar code consists of bars, etc. This aversion to internal plurals has exceptions, most of which are amenable to explanations of various types (see Dierickx 1970, Johansson 1980, Jamet 2008). The situation in French is very different as there do not seem to be any restrictions on N2, the non-head being in the plural when it has a multiple denotation, as in *base-travaux°* 'base works = engineering base', *bâtiment voyageurs* 'building travellers = passenger building' and *code-barres* 'code bars = barcode'.[18]

Other points of comparison exist. Recursivity in the formation of subordinative NNN units, which was mentioned in section 2, is much more frequent in English than in French, where the number of examples we have been able to collect is limited, but non-null, *contra* Delfitto & Melloni's (2009) claim that recursion in Romance is limited to coordinative units. Among these French units, the direction of branching is unclear in examples like *officier sécurité incendie* 'officer safety fire = fire safety officer' and *plan épargne logement* 'plan savings housing = housing savings account', where N1N2 and N2N3 are separately lexicalized. These NNN units might result from a blending-like process where one copy of the constituent common to the two source composite units is retained, in which case these examples would not be legitimate cases of recursivity. Note, however, that in *compte chèque-cadeau* 'account cheque gift = gift-cheque account', where N1N2 and N2N3 are also separately lexicalized, the structure $[N[NN]]_N$ is clear. The direction of branching is unclear again in *contrat initiative emploi* 'contract initiative employment = a type of government-subsidized employment contract', in which no pre-lexicalized NN sequence appears and the semantic hierarchy is ambiguous. In contrast, the branching in English NNN compounds is almost always clear (the percentages in our database of 295 English NNN items are 92 for left-branching units, and 7 for right-branching ones, while only 1% are undecidable). *Cigarette bout filtre* is right-branching, and units like *crédit impôt recherche*, or *poisson-coffre pintade* 'fish coffer Guinea-fowl = white-spotted boxfish' and *poisson-ange amiral* 'fish angel admiral = blue-girdled angelfish', which are two attributives, are undisputably left-branching, which contradicts the claim by Haider (2001) that left-headed compounding cannot be recursive (note that these examples

cannot be interpreted as calques from English). There remains the fact, however, that recursivity is very limited in French.

A further point of comparison lies in the fact that the link between non-head and head in French units may be rather loose, which appears in occurrences such as the following:

(49)   *base vie, logistique et de soutien*

      'base life, logistic and of support'

      living quarters cum logistics and support base.

Base vie may be lexicalized, but here *vie* freely coordinates in a clearly syntactic fashion with an adjective and a prepositional phrase. Adjectival modifiers can also compete with nominal ones for proximity to the head in cases of multiple modification:

(50)   *presse écrite magazine et quotidienne*

      'press written magazine and daily'

      paper dailies and magazines

(51)   *prix net vendeur*

      'price net seller'

      seller's net price.

A number of facts can be further adduced to describe the status of French subordinatives:

    – although relational subordinatives have been present in the language since the 13th century, they were rare before 1850 (Arnaud 2003: 119-31) and their numbers increased considerably in the late 19th and the 20th centuries, while NNs have been present since the earliest English documents;

    – many units seem to be calque translations of English compounds, for example *price competitiveness ~ compétitivité prix, tunnel current ~ courant tunnel, project team ~ équipe projet* and *health food ~ aliment santé*; however, as just mentioned, the word-formation pattern was present in the language at a time that rules out initial, structural influence from English;

    – some new formations like *radar tronçon* 'radar road-section = a speed radar operating over a stretch of road' may sound strange to some informants in the same way as earlier examples of

what he called '*composition avec le génitif*', that is our relationals, did to Darmesteter (1874: 240),[19] and a proper name like *la Ferme Célébrités* 'the Farm Celebrities = Celebrities' Farm' (a 2004-2005 French reality television show) is often judged strange;

– in spite of their relative productivity, they still are a minority when synchronically compared with prepositional units like *communauté d'agglomération* 'community of conurbation = a French metropolitan government structure', *puits de carbone* 'pit of carbon = carbon sink', *obus à fléchettes* 'shell at / to flechettes = flechette shell' and *puce à ADN* 'chip at / to DNA = DNA microarray', all recent examples;

– a prepositional variant exists in many cases, as in *village vacances ~ village de vacances* 'village (of) holidays = holiday village', *point presse ~ point de presse* 'spot (of) press = press conference' and *zone vie ~ zone de vie* 'zone (of) life = life zone' (we return to this point later);

– some of the non-head-to-head relations of English relationals seem to be missing in our database of French units, such as the HABITAT locative relation found in *sea otter* or *watercress*, although this may simply be due to the smaller number of French units and it may just be a question of time before they appear. Alternatively, it may be that, as suggested by Bauer (2009), not all languages exhibit the semantic variability of Germanic compounds, and so French is limited in this respect.

The picture which emerges is that subordinative $[NN]_N$ units are less central in the word-formation devices of French than they are in English and that some of their characteristics are rather different.

As announced in the Introduction, we have so far used the terms $[NN]_N$ UNIT and COMPOSITE for convenience, and in this we have distinguished ourselves from a French tradition (for example Guilbert 1971, Mathieu-Colas 1996, Gross 1996: 33, Apothéloz 2002: 18-19 and Riegel *et al*. 2009: 912), where the term MOT COMPOSÉ is extended to lexicalized multiword constructions, including $[NAdj]_N$ and $[N$ prep. $N]_N$. In theoretical frameworks like Generative Grammar, where morphology and syntax are separate modules, however, the question of the grammatical status of French $[NN]_N$ units is an important one.

In some lexicalist approaches, like Di Sciullo & Williams' (1987), the Right-Hand Head Rule (Williams 1981: 61) leads to the radical consequence that French is seen as entirely devoid of compounds (Di Sciullo & Williams 1987: 83), and relationals are relegated to the status of listed 'phrasal idioms' (1987: 81). Such positions were to be opposed at a later stage: ten Hacken (1999) considers that the Right-Hand Head Rule arbitrarily separates left-headed constructions from their right-headed counterparts in other languages, and so should not be part of a cross-linguistic definition of compounds, and, more recently, Guevara & Scalise (2009: 122) reject the claim that compounds can only be right-headed as 'just a theory-internal construct'.

Considering the French lexicon, Corbin (1992) reserves the label COMPOSÉ for sequences that cannot be constructed by the syntax and shows that this covers the case of the majority of NNs, at least those where N1 is the hyperonym of the NN unit. In Corbin (1997), she criticizes Zwanenburg (1992), who, following Di Sciullo & Williams' position, compares *homme grenouille* 'man frog = frogman' and *avocat ami* 'lawyer friend', or *timbre poste* and *projet Delors* 'plan Delors = Delors plan'. According to Zwanenburg, as the second members of those pairs of NN sequences are clearly syntactic, the first ones must also be, but Corbin remarks that the non-head-to-head relations are very different: in *avocat ami* it is a qualifying one, while it has a categorizing effect in *homme grenouille*; similarly, while the proper name *Delors* serves to identify one particular plan, *poste* forms the name of a subcategory of *timbre*. She concludes that *homme-grenouille* and *timbre-poste* have semantic properties different from those of syntactic NN structures, and are therefore compound lexemes.

Fradin (2009), who, like Corbin, applies the principle that what may be formed by syntax should not be attributed to morphology, notes that it is not always easy to determine which French NN constructions are actually morphological outputs. Like Corbin (1992), he views coordinative units as compounds, and, with respect to subordinative units, he considers that no arguments justify a rejection of the compound status of analogicals, but sees relationals in a different light. He claims first that '[…] N2 never introduces a semantic predicate that N1 would be an argument of' (Fradin 2009: 432). If we interpret this correctly, a unit like (12), repeated here, is however a counter-example:

(12)   *équipe action*

'team action' = action team

act (team).

Fradin (2009: 432-3) also notes with the examples *responsable surgelés / jouets / charcuterie* 'person-in-charge frozen foods / toys / delicatessen = frozen foods / toys / delicatessen manager' that:

> […] these constructs usually denote entities that belong to a conceptual or technological classificatory system and N2 provides a clue about the place the construct's referent occupies within the classification in question.

The adverb 'usually', however, leaves room for [NN]$_N$ units denoting entities that do not correspond to this description, such as *gauche caviar* 'left caviare = champagne left', *homme canon* 'man gun = human cannonball', *moto-crottes* and *consultation mémoire* 'medical-consultation memory = a consultation on memory problems'. Third, Fradin (2009: 433) notes that all his examples '[…] also exist with a preposition heading a PP including either an N that can be bare [...] or an (incomplete) NP':

(52)   (a)   *impression (par) laser*

'printing (by) laser'

(b)   *roman (avec (des)) photos*

'novel (with (ø)) photos'

(c)   *langage (de l') auteur*

'language (of the) author'

Although the author does not say so explicitly, this would point out to the presence of a prepositional phrase in the formation of the [NN]$_N$ units, which echoes earlier attempts at classifying relations in French [NN]$_N$ units in terms of an ellipted preposition (Grieve-Schumacher 1960: 30, Haensch & Lallemand-Rietkötter 1972: 59). One can object, however, that the few corpus occurrences of *roman avec des photos* 'novel with photos = a text illustrated with photos' clearly refer to something different from *roman-photo* ('novel photo = a photonovel, that is a narrative consisting of a succession of photos with captions or bubbles'). Another of Fradin's examples, *impôt sécheresse*, has only one reliable Web occurrence in its prepositional version.[20] In fact, French relationals may exist without a prepositional variant or with that variant only attested in a very tiny minority of occurrences. This is the case, for

instance, of *avion suicide* 'plane suicide = suicide plane', *doublure lumière* 'double light = an actor's stand-in', *enfant bulle* 'child bubble = bubble child', *papier carbone* 'paper carbon = carbon paper', *sucre glace* 'sugar ice = icing sugar', *wagon lits* 'carriage beds = sleeping car' *micro trottoir*, *pierre papier*, *pont canal* and *timbre poste*. Finally, Fradin suggests that the fact that the numbers of N1 and N2 may be different, as in *roman photos*, is similar to the situation in prepositional units and might be explained by a syntactic origin. Another explanation seems possible: in an attributive subordinative compound, analogy tends to be one-to-one (in *crocodile clip*, a clip is compared to a crocodile), which would explain why N2 is in the singular when N1 is too, while in relationals, the merological relation, for instance, may logically be one-to-several, as in *code-barres*. All in all, Fradin's arguments against the morphological status of relational units do not close the debate.

From an overall examination of problems with the Lexical Integrity Hypothesis in various languages, Lieber & Scalise (2007) conclude that languages can allow morphology to use syntactic phrases, and that phrases can receive word status during the formation of the complex units that include them. Referring to facts similar to those presented above in this section for French units with phrasal constituents (Italian examples where, in the input, N1 is a morphological word and the non-head is a product of syntax, while the whole unit behaves as a morphological word), Montermini (2008) and Arcodia *et al*. (2009) see in them a serious challenge to modular models of grammar. Like Lieber & Scalise (2007), they conclude that syntax can systematically, and not exceptionally, as believed before, feed into morphology.

One approach that takes this fact into account is Di Sciullo's (2005) model, in which morphology and syntax operate in parallel and can exchange sequences: while English compounds are derived in the morphological space, French units are derived in the syntactic space and then transferred to the morphological space. Note that here compounds of all categories result from the same derivation rules. In particular, since French units are derived syntactically, they must include a functional projection realized as an operator providing the semantic relation between the two nouns. In the case of coordinatives, this operator is AND / OR and in the case of relationals, it is SORT, which can either surface in prepositional units (*lampe de chevet* 'lamp of bedside = bedside lamp') or not be legible at the phonological interface (*lampe-tempête* 'lamp storm = storm lantern'; our examples). This is an appealing solution, but problems

with this 'theory-internal construct' are the *ad hoc* character and semantic indetermination of SORT and the fact that the model includes similar derivation mechanisms for subordinatives and coordinatives in spite of their crucial differences.

7 Conclusion

Germanic languages make more use of composite lexical units than Romance languages. Our classification, however, shows that for every class of English [NN]$_N$ compounds it is possible to find corresponding French examples. Some differences in the morphology-syntax interface are particularly noteworthy, however: while English units have external plurals, French ones take the plural flexion on their heads or on both constituents, so structure-internally; they also display less internal cohesion. From a typological perspective, they must be considered as less prototypical compounds than English units.

[1] We would like to thank two anonymous referees whose comments and suggestions helped us improve the original manuscript.

[2] 'French, in fact, possesses diverse kinds of composition, but is seemingly afraid to use them' (our translation).

[3] There is considerable variation in the spelling of $[NN]_N$ lexical units and we have generally adopted the spelling which is dominant in the dictionaries or corpora we have examined.

[4] Among the exceptions in our data are *troutperch*, which is left-stressed, and *oxidation-reduction*, *toxin-antitoxin* and *tractor trailer*, which are double-stressed.

[5] For extensive surveys of French NN sequences, see Noailly (1990) and Gushchina-Magno (2009).

[6] In T6, the term HYBRID is used in a non-technical sense. It may refer to perceptual hybridity only, as in the case of *troutperch* (*Percopsis omiscomaycus*).

[7] A North-American term which corresponds to British *articulated lorry*.

[8] Marchand (1969: 40) calls them SUBSUMPTIVE COMPOUNDS; Dressler's (2006) term is PLEONASTIC COMPOUNDS. Scalise & Bisetto (2009) classify *oak tree* as a coordinative, but linear irreversibility (**tree oak*) demonstrates that this unit has a single head: *tree*.

[9] *Child prodigy* is an exception. Due in part to their transparency, units of this kind are virtually always absent from dictionaries.

[10] For an extensive discussion of intermediate examples in French, see Gushchina-Magno (2009: 285-310).

[11] See Bauer (1998), Olsen (2000) and Giegerich (2004, 2009); the research of Plag and his colleagues (Plag 2006, 2010, Plag *et al.* 2008, Arndt-Lappe 2011) points at an emergent, non-rule-governed nature of English compound stress based on constituent families.

[12] The following example, which admittedly involves a [VN]$_N$ unit, was heard on television; there was no recoverable antecedent earlier in the discourse: *Des pare-feu$_i$ seront établis pour limiter sa$_i$ propagation* 'fire$_i$-breaks will be established to limit its$_i$ propagation'.

[13] Plurals appear in the spelling of the vast majority of French nouns, but the only plural forms that are pronounced differently from the singulars are those of irregular nouns. Referring to pluralization in a discussion of French composites is therefore not without its problems, as subordinative relational units with irregular plurals can be counted on the fingers of one hand, and, due to the non-pronunciation of regular plurals, spelling fluctuates widely. This probably explains why Rosenberg (2007) found a tendency to double plurals in relational as well as attributive subordinatives in her corpus. It is therefore interesting to turn to prescriptive sources, although composites are not always dealt with adequately in dictionaries (for example, the plural of *bière pression* 'beer pressure = draught beer' is absent from both Imbs 2002 and Rey-Debove & Rey 2011).

[14] The symbol '°' indicates that number is audible.

[15] *Ponts-canaux°* 'bridges canals' is also found.

[16] See also the discussion of Breton loose compounds in Stump (2005).

[17] They behave differently from Italian attributives, which are pluralized on their heads only, and Spanish ones, which admit single and double plural variants (Guevara & Scalise 2009).

[18] We also have one example of an attributive with a sizable minority of occurrences with a plural non-head: *coco-fesse(s)* 'coconut buttock(s) = sea coconut' (but see Note 13).

[19] 'Les rares formations qu'on lui doit ont une apparence singulière, étrange, qui étonne dès l'abord ; elles semblent dépaysées au milieu de la langue.' ('The few formations that are due to it have a strange, singular appearance, which causes surprise at first sight; they seem out of place amidst the language', our translation)

[20] 'si on ne veut pas voir un impot [sic] de sécheresse' ('if you don't want to see a tax of drought'). Retrieved 20/12/2011 from http://www.laboutiquedecasawan.blogspot.com/2011/05/de-leau-on-en-veut.html.

References

Anderson, Stephen R. 1992. *A-morphous morphology*. Cambridge: Cambridge University Press.

Apothéloz, Denis 2002. *La construction du lexique français: principes de morphologie dérivationnelle*. Gap: Ophrys.

Arcodia, Giorgio F., Nicola Grandi & Fabio Montermini 2009. Hierarchical NN compounds in a cross-linguistic perspective. *Italian Journal of Linguistics / Rivista di Linguistica* 21(1): 11-33.

Arnaud, Pierre J.L. 2003. *Les composés* timbre-poste. Lyon: Presses Universitaires de Lyon.

Arnaud, Pierre J.L. 2010. Judging the degree of adjectivization of English nouns: An investigation of the discourse behaviour of a sample of frequent premodifier nouns. *English Studies* 91(3): 303-327.

Arndt-Lappe, Sabine 2011. Towards an exemplar-based model of stress in English noun-noun compounds. *Journal of Linguistics* 47(3): 549-585.

Baroni, Marco, Emiliano Guevara & Vito Pirrelli 2007. NN compounds in Italian: modelling category induction and analogical extension. *Lingue e Linguaggio* 6(2): 263-290.

Bauer, Laurie 1998. When is a sequence of two nouns a compound in English? *English Language and Linguistics* 2(1): 65-86.

Bauer, Laurie 2008a. Les composés exocentriques de l'anglais. In Dany Amiot (ed.), *La composition dans une perspective typologique*. Arras: Artois Presses Université. 35-47.

Bauer, Laurie 2008b. Exocentric compounds. *Morphology* 18(1): 51-74.

Bauer, Laurie 2009. Typology of compounds. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of compounding*. Oxford: Oxford University Press. 343-356.

Becker, Thomas 1992. Compounding in German. *Italian Journal of Linguistics / Rivista di Linguistica* 4(1): 5-36.

Benczes, Réka 2006. *Creative compounding in English: the semantics of metaphorical and metonymical noun-noun combinations*. Amsterdam and Philadelphia: Benjamins.

Bisetto, Antonietta & Sergio Scalise 2005. The classification of compounds. *Lingue e Linguaggio* 4(2): 319-332.

Bloomfield, Leonard 1933. *Language*. New York: Holt.

Booij, Geert 2007 [2005]. *The grammar of words: An introduction to linguistic morphology*. 2nd edn. Oxford: Oxford University Press.

Booij, Geert 2010. *Construction morphology*. Oxford: Oxford University Press.

Cambridge advanced learners' dictionary online. Electronic site: http://dictionary.cambridge.org/ dictionary/british/ (last accessed 16/02/2012).

Corbin, Danielle 1992. Hypothèses sur les frontières de la composition nominale. *Cahiers de Grammaire* 17: 27-55.

Corbin, Danielle 1997. Locutions, composés, unités polylexématiques: lexicalisation et mode de construction. In Michel Martins-Baltar (ed.), *La locution entre langue et usages*. Fontenay-aux-Roses: ENS Éditions Fontenay / Saint-Cloud. 53-101.

Cruse, D. A. 1986. *Lexical semantics*. Cambridge: Cambridge University Press.

Darmesteter, Arsène 1874. *Traité de la formation des mots composés dans la langue française comparée aux autres langues romanes et au latin*. Paris: Franck.

Darmesteter, Arsène 1877. *De la création actuelle des mots nouveaux dans la langue française et des lois qui la régissent*. Paris: Vieweg.

Delfitto, Denis & Chiara Melloni 2009. Compounds don't come easy. *Lingue e Linguaggio* 8(1): 75-104.

Di Sciullo, Anna-Maria 2005. Decomposing compounds. *SKASE Journal of Theoretical Linguistics* 2(3): 14-33.

Di Sciullo, Anna-Maria & Edwin Williams 1987. *On the definition of word*. Cambridge MA: The MIT Press.

Dierickx, Jean 1970. Why are plural attributives becoming more common? In Jean Dierickx & Yvan Lebrun (eds.), *Linguistique contemporaine: hommage à Eric Buyssens*. Bruxelles: Éditions de l'institut de sociologie. 39-46.

Downing, Pamela 1977. On the creation and use of English compound nouns. *Language* 53(4): 810-842.

Dressler, Wolfgang U. 2006. Compound types. In Gary Libben & Gonia Jarema (eds.), *The representation and processing of compound words*. Oxford: Oxford University Press. 23-44.

Fabb, Nigel 1998. Compounding. In Andrew Spencer & Arnold M. Zwicky (eds.), *The handbook of morphology*. Oxford: Blackwell. 66-83.

Fradin, Bernard 2009. IE, Romance: French. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of compounding*. Oxford: Oxford University Press. 417-435.

Gaeta, Livio & Davide Ricca 2009. *Composita solvantur*: compounds as lexical units or morphological objects? *Italian Journal of Linguistics / Rivista di Linguistica* 21(1): 35-70.

Giegerich, Heinz J. 2004. Compound or phrase? English noun-plus-noun constructions and the stress criterion. *English Language and Linguistics* 8(1): 1-24.

Giegerich, Heinz J. 2009. The English compound stress myth. *Word Structure* 2(1): 1-17.

Grieve-Schumacher, Madeleine 1960. *Die Nominalkomposition im Französischen*. Winterthur: Keller.

Gross, Gaston 1996. *Les expressions figées en français. Noms composés et autres locutions*. Gap: Ophrys.

Guevara, Emiliano & Sergio Scalise 2009. Searching for universals in compounding. In Sergio Scalise, Elisabetta Magni & Antonietta Bisetto (eds.), *Universals of language today*. Dordrecht: Springer. 101-128.

Guilbert, Louis 1971. De la formation des unités lexicales. In Louis Guilbert, René Lagane & Georges Niobey (eds.), *Grand Larousse de la langue française*. Paris: Larousse. ix-lxxxi.

Gushchina-Magno, Olga 2009. *Les constructions '(article) + substantif + substantif' en français moderne*. Unpublished doctoral dissertation. Université Paris-IV Sorbonne.

ten Hacken, Pius 1999. Motivated tests for compounding. *Acta Linguistica Hafniensia* 31: 27-58.

Haider, Hubert 2001. Why are there no complex head-initial compounds? In Chris Schaner-Wolles, John R. Rennison & Friedrich Neubart (eds.), *Naturally! Linguistic studies in honour of Wolfgang Ulrich Dressler presented on the occasion of his 60th birthday*. Turin: Rosenberg and Sellier. 165-174.

Haensch, Günter & Annette Lallemand-Rietkötter 1972. *Wortbildungslehre des modernen Französisch*. Munich: Hueber.

Haspelmath, Martin & Andrea D. Sims 2010 [2002]. *Understanding morphology*. 2nd edn. London: Hodder Education.

Hatcher, Anna Granville 1951. *Modern English word-formation and neo-Latin. A study of the origins of English (French, Italian, German) copulative compounds*. Baltimore MD: The Johns Hopkins Press.

Hatcher, Anna Granville 1960. An introduction to the analysis of English noun compounds. *Word* 16(3): 356-373.

Imbs, Paul (ed.) 2002. *Trésor de la langue française informatisé*, version 4. Nancy: ATILF. Electronic site: http://atilf.atilf.fr/tlfv4.htm (last accessed 01/12/2011).

Jackendoff, Ray 2009. Compounding in the parallel architecture and conceptual semantics. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of compounding*. Oxford: Oxford University Press. 105-128.

Jamet, Denis 2008. Les *plural attributives* en anglais néo-zélandais. In Denis Jamet (ed.), *Mélanges en hommage à Malcolm Clay*. Lyon: Université Jean Moulin. 29-57.

Johansson, Stig 1980. *Plural attributive nouns in present-day English*. Lund: CWK Gleerup.

Kim, Gyung-Ran 2001. Obstruent alternations in sub-compounds. *Studies in Phonetics, Phonology and Morphology* 7(2): 315-331.

Lees, Robert B. 1970. Problems in the analysis of English nominal compounds. In Manfred Bierwisch & Karl Erich Heidolph (eds.), *Progress in linguistics*. The Hague: Mouton.

Levi, Judith N. 1978. *The syntax and semantics of complex nominals*. New York: Academic Press.

Lieber, Rochelle 2004. *Morphology and lexical semantics*. Cambridge: Cambridge University Press.

Lieber, Rochelle 2009. A lexical approach to compounding. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of compounding*. Oxford: Oxford University Press. 78-104.

Lieber, Rochelle & Sergio Scalise 2007. The Lexical Integrity Hypothesis in a new theoretical universe. In Geert Booij, Luca Ducceschi, Bernard Fradin, Emiliano Guevara, Angela Ralli & Sergio Scalise (eds.), *On-line proceedings of the Fifth Mediterranean Morphology Meeting (MMM5)*. Retrieved 01/12/2011 from http://mmm.lingue.unibo.it/proc-mmm5.php.

Lieber, Rochelle & Pavol Štekauer (eds.) 2009. *The Oxford handbook of compounding*. Oxford: Oxford University Press.

Longman dictionary of contemporary English online. Electronic site: http://www.ldoceonline.com (last accessed 01/12/2011).

Macmillan dictionary online. Electronic site: http://www.macmillandictionary.com/ (last accessed 01/12/2011).

Marchand, Hans 1969 [1960]. *The categories and types of present-day English word-formation. A synchronic-diachronic approach*. 2nd edn. Munich: Beck.

Mathieu-Colas, Michel 1996. Essai de typologie des noms composés français. *Cahiers de Lexicologie* 69(2): 71-125.

Mätzner, Eduard 1860. *Englische Grammatik: die Lehre vom Worte*. Berlin: Weidmann.

Montermini, Fabio 2008. La composition en italien dans un cadre de morphologie lexématique, In Dany Amiot (ed.), *La composition dans une perspective typologique.* Arras: Artois Presses Université. 161-187.

Neef, Martin 2009. IE, Germanic: German. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of compounding*. Oxford: Oxford University Press. 386-399.

Noailly, Michèle 1990. *Le substantif épithète*. Paris: Presses Universitaires de France.

Noordegraaf, Jan 1989. From the history of the term 'exocentric'. *Historiographia Linguistica* 16(1-2): 211-216.

Olsen, Susan 2000. Compounding and stress in English: a closer look at the boundary between morphology and syntax. *Linguistische Berichte* 181: 55-69.

Olsen, Susan 2001. Copulative compounds: a closer look at the interface between syntax and morphology, In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 2000*. Dordrecht: Kluwer Academic Publishers. 279-320.

Olsen, Susan 2004. Coordination in morphology and syntax: the case of copulative compounds. In Alice ter Meulen & Werner Abraham (eds.), *The composition of meaning: from lexeme to discourse*. Amsterdam and Philadelphia: Benjamins. 17-37.

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

Plag, Ingo 2006. The variability of compound stress in English: structural, semantic and analogical factors. *English Language and Linguistics* 10(1): 143-172.

Plag, Ingo 2010. Compound stress assignment by analogy: the constituent family bias. *Zeitschrift für Sprachwissenschaft* 29(2): 243-282.

Plag, Ingo, Gero Kunter, Sabine Lappe & Maria Braun 2008. The role of semantics, argument structure and lexicalisation in compound stress assignment in English. *Language* 84(4): 760-794.

Rey-Debove, Josette & Alain Rey (eds.) 2011. *Le petit Robert de la langue française 2012*. Paris: Le Robert.

Riegel, Martin 1988. Vrais et faux noms composés: Les séquences binomiales en français moderne. In Dominique Weil & Huguette Fugier (eds.), *Actes du Troisième Colloque Régional de Linguistique, Strasbourg, 28-29 avril 1988*. Strasbourg: Université des Sciences Humaines. 371-394.

Riegel, Martin 1991. Ces noms dits composés: arguments et critères. *Studia Romanica Posnaniensia* 16: 148-161.

Riegel, Martin, Jean-Christophe Pellat & René Rioul 2009 [1994]. *Grammaire méthodique du français*. 4th edn. Paris: Presses Universitaires de France.

Rosenberg, Maria 2007. Classification, headedness and pluralization: corpus evidence from French compounds. *Acta Linguistica Hungarica* 54(3): 341-360.

Scalise, Sergio & Antonietta Bisetto 2009. The classification of compounds. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of compounding*. Oxford: Oxford University Press. 34-53.

Scalise, Sergio & Antonio Fábregas 2010. The head in compounding. In Sergio Scalise & Irene Vogel (eds.), *Cross-disciplinary issues in compounding*. Amsterdam and Philadelphia: Benjamins. 109-125.

Scalise, Sergio & Irene Vogel (eds.) 2010. *Cross-disciplinary issues in compounding*. Amsterdam and Philadelphia: Benjamins.

Selkirk, Elisabeth O. (1982). *The syntax of words*. Cambridge MA: MIT Press.

Soanes, Catherine & Angus Stevenson (eds.) 2004 [1911]. *Concise Oxford English dictionary*. 11th edn. Oxford: Oxford University Press.

Stump, Gregory T. 2005. Word-formation and inflectional morphology. In Pavol Štekauer & Rochelle Lieber (eds.), *Handbook of word-formation*. Dordrecht: Springer. 49-71.

Ward, Gregory, Richard Sproat & Gail McKoon 1991. A pragmatic analysis of so-called anaphoric islands. *Language* 67(3): 439-474.

Williams, Edwin 1981. On the notions 'lexically related' and 'head of a word'. *Linguistic Inquiry* 12(2): 245-274.

WordNet, version 3.1 Electronic site: http://wordnetweb.princeton.edu/perl/webwn (last accessed 01/12/2011).

Zwanenburg, Wiecher 1992. Compounding in French. *Italian Journal of Linguistics / Rivista di Linguistica* 4(1): 221-240.

Author's address: *(Pierre J.L. Arnaud & Vincent Renner)*

Centre de Recherche en Terminologie et Traduction

Université Lumière Lyon 2

69365 Lyon CEDEX 07

France

E-mail: pierre.arnaud@univ-lyon2.fr, vincent.renner@univ-lyon2.fr