Lecture Notes

LEXICAL MATTERS

edited by
Ivan A. Sag and Anna Szabolcsi

CSLI
CENTER FOR THE STUDY
OF LANGUAGE
AND INFORMATION

6

The Stress and Structure of Modified Noun Phrases in English

MARK LIBERMAN AND RICHARD SPROAT

1 Introduction

Our topic is the stress pattern of English noun phrases in which the head noun is preceded by a sequence of modifiers.[1] We assume a context of use that is rhetorically stress-neutral; the phenomena of FOCUS, CONTRAST and ANAPHORA—henceforth FCA—are taken to be perturbations of the patterns that we discuss. We attempt to establish the basic regularities that shape the complex data in this area, against the background of a broad (and thus complex) description. Our purpose is to establish an adequate set of descriptive categories, able to support a formal model of the syntax, semantics and prosody of complex nominals. We would like such a model to be adequate for parsing and assigning stress to modified noun phrases in unconstrained English text. We start with a careful description of the phenomena, followed by a more formal account of the proposed syntactic analysis, and a sketch of the implications for parsing and stress-assignment algorithms.

Many syntacticians (e.g., Jackendoff 1977) have noted the existence of at least four distinct prenominal positions, arranged in a right-branching structure:

(1)

4	3	2	1	0
the	three	exotic	chess	boards
John's	many	large	book	bags
those	few	Chinese	store	owners

Position 4 is stereotypically occupied by articles, demonstratives, and possessive phrases; position 3 by certain quantifiers and numerals; position

[1]The authors would like to thank an anonymous reviewer for useful comments, and Julia Hirschberg and Mats Rooth for some discussion.

Lexical Matters. Ivan A. Sag and Anna Szabolcsi, eds.

2 by adjectives; and position 1 by nouns making up the initial element of a compound word. We will refer to items in positions 1 and 2 as MODIFIERS.

Each of the positions in (1) may be occupied by an item that has internal structure of its own: *John's brother's* (position 1), *distressingly few* (3), *more exotic* (2), *liquor store* (1). Items characteristic of positions 1 and 2 are often repeated, either in parallel or a layered fashion: *powerful, luxurious automobiles*; *powerful economic forces*. Normally, a strict ordering of these positions is required—phrases like **large many John's book bags* are about as wrong as arrangements of English words can be.

The lefthand edge of English noun phrases is more complex than the simple pattern (1) indicates. For instance, certain quantifiers cannot substitute in position 3 of the pattern, co-occurring with articles, demonstratives and possessive phrases only in so-called *partitive* constructions: **John's all/some/any/each large book bags*, *all/some/any/each of John's large book bags*; see Jackendoff 1977, pages 104ff for some discussion. Other complexities arise in the handling of pronouns, *such*, definiteness, and so forth. However, we will not discuss the intricacies of material to the left of position 2 in (1); our concern is with the complexities of the inner structure, and their influence on stress patterns.

1.1 A Sketch of Our Conclusions

We take a traditional view of modifiers in positions 1 and 2 of (1): a position 1 modifier combines with a noun to form a compound noun, whereas a position 2 modifier forms a phrasal category. Expressed in terms of X-bar theory, position 1 is filled by modifiers of N^0 whereas position 2 contains modifiers of N^1. The structure of *large exotic chess boards* is thus:

(2) $[_{N^1}$ large $[_{N^1}$ exotic $[_{N^0}$ chess boards]]]

Position 1 modifiers—COMPOUND MODIFIERS—are thus adjunctions to N^0, whereas position 2 modifiers—PHRASAL MODIFIERS—are daughters of N^1, modifying a right sister that is either N^1 or N^0.[2]

We shall agree with the traditional generative view (Chomsky and Halle 1968, Liberman and Prince 1977, Hayes 1980) that constructions involving positions 1 and 2 in (1) are assigned different default stress patterns due to the difference in category of their parent node. English stress is normally assigned recursively to rightmost elements, but the stress rule for nouns, simple or compound, will ignore a single non-complex element at each level. This implies that N^1 constructions are "right dominant"—i.e., have main stress on the head noun—while N^0 constructions are "left dominant"—i.e., have main stress on the modifier—as long as the head is

[2]We know of no distinction in types of modification that would motivate maintaining these two alternative categories for the right sister of position 2 modifiers, so one could assume that such modifiers are always adjoined to N^1. Under that assumption, the sister of *exotic* in (2) would be more correctly given as $[_{N^1}$ $[_{N^0}$ chess boards]].

a single word. We thus disagree with the view (Bolinger 1972, Ladd 1984) that the difference between compound and phrasal stress is derivable from FCA effects.

Position 1 modifiers are usually taken to be joined with their heads in a separate lexical or morphological component, whereas position 2 modifiers are taken to be syntactic in origin; a recent expression of this view is given in Di Sciullo and Williams 1987. We agree that constructions dominated by N^0 are words, but we see no good evidence for assuming that they should be considered anything other than syntactic constructions in English. We shall return in a later section to a fuller discussion of the theoretical issues in English noun phrase modification.[3]

We also agree with the traditional view that the two different types of syntactic modification imply different sorts of semantic relationship as well. However, as will become clear from the discussion below, there does not seem to be a single, clean semantic distinction such that all nominals with lefthand stress will fall into one semantic class and all nominals with righthand stress will fall into the other.

The main novelty of our position is this: we argue that both position 1 and position 2 may be occupied by modifiers of a wide variety of categories. We show in particular that positions 1 and 2 may be filled by both adjectives and nouns or phrases made up of these and other categories, sometimes linked with the possessive *'s*. This implies that for any modifier-noun sequence, both N^0 and N^1 structures are always available in principle. Thus the contextually appropriate parsing of such phrases may require a judgment as to the relative plausibility of the semantic relations implied by the structural choices. This is at least somewhat consistent with Bolinger's (1972) dictum that "accent is predictable (if you're a mind reader)". However, we will suggest that there are some strategies that can achieve fairly high accuracy without telepathic assistance.

1.2 The Problem

It is common in running text or speech in English to find PREMODIFIED NOMINALS containing one or more modifiers, such as the examples in (1). Depending on the style of the material and on the definition of 'phrase', something like 30 to 70 percent of all phrases can be expected to end in such units. As a matter of practice, the location of the main stress in these sequences is quite variable. Although the final noun is the commonest location, pre-final main stress is also quite frequent. Depending on the type of material studied, somewhere between 10 and 60 percent of the premodi-

[3]For those readers familiar with the work of Abney 1987, we note that we shall use the terms 'noun phrase' and 'NP' in their traditional sense for the bulk of this paper. Also, we shall generally use 'XP' (e.g., NP, AP,...) to refer to the (contextually) maximal projection of the category in question with no commitment to the bar level of this projection.

fied noun constituents show pre-final main stress, with 30-40 percent being typical of newspaper writing.

There are several sorts of reasons why the main stress in a noun group might fall before the head. We distinguish cases where the preterminal stress serves to underline an important word, or to avoid stressing a redundant one, from cases where a preterminal stress pattern is normal or natural for the phrase in neutral contexts of use. In the following example, and henceforth in this paper, boldface is used, where needed, to mark the word that bears the main stress:

(3) a. Stress pattern "natural" for the phrase:
   i. At 9:00, there is an important **meeting**.
   ii. At 9:00, there is a **staff** meeting.
   iii. John was wearing a red **jacket**.
   iv. John was wearing a **life** jacket.

   b. Stress pattern determined by "FCA" considerations
   i. We're only concerned with **solvable** problems.
   ii. He replaced his **low-interest** bonds with **high-interest** bonds.

While FCA effects are not uncommon and cannot be ignored, it is nonetheless true that FCA phenomena must in turn interact with other principles—such as the lexical stress pattern of polysyllabic words—to produce observed stress patterns; see Hirschberg 1990 for some recent discussion of issues in modeling some FCA effects and their interaction with lexical stress placement. In particular, a familiar generalization seems to underlie the stress pattern of cases like those in (3a), where $[A\ N]$ phrases typically show main stress on the head noun (righthand stress), while $[N\ N]$ units typically show stress on the noun in the lefthand position. These simple generalizations hold true more often than not. We can provide a rough experimental check by having someone read some text, and counting up the stress patterns employed for constituents of the appropriate kinds. In our experience, more than 90% of $[A\ N]$ units will be read with righthand main stress, while around 75% of $[N\ N]$ units will show lefthand main stress. .

Although the contrary cases remain fairly frequent, we might suppose that they represent the expected effect of the phenomena of FCA. Because theories of FCA phenomena are not very well developed, it is hard to check this notion with total assurance. However, one typically has some feeling for the application of such analyses in particular cases, and we can look for them in all the textual examples whose stress patterns are contrary to the predictions of part of speech sequence. On this basis, FCA-type explanations do not seem to help very much, especially for the $[N\ N]$ anomalies.

The exact statistics depend very much on the style of the material

surveyed, but two sample cases will give the flavor of the situation. In one text (taken from a book on computer vision), 190 out of 214 [*A N*] constituents showed righthand stress, while 76 out of 92 [*N N*] constituents showed lefthand stress.[4] A plausible explanation in terms of FCA could be found for 18 of the 24 left-stressed [*A N*] units, and for none of the 16 right-stressed [*N N*] units. Thus we are left with about 3% anomalous left-stressed [*A N*] expressions and about 17% anomalous right-stressed [*N N*] expressions. In a second text (several stories from the main section of the *New York Times*), 153 out of 169 [*A N*] constituents had the expected righthand stress, while 102 out of 138 [*N N*] constituents had the expected lefthand stress. FCA-type explanations applied to 8 of the 16 left-stressed [*A N*] units, and to none of the 36 [*N N*] units, leaving about 5% non-FCA left-stressed [*A N*] expressions, and about 26% non-FCA right-stressed [*N N*] expressions. In these two sample texts, only 60% of the "anomalous" [*A N*] cases (and none of the anomalous [*N N*] cases) had a plausible FCA explanation.[5] This means that in 6.7% of all [*A N*] sequences there was a useful FCA explanation for the stress pattern (be it left or right), and in none of the [*N N*] cases.

So, while the stress pattern of an [*X N*] expression is strongly correlated with the lexical category of the word preceding the head noun, there are quite a few apparent exceptions, most of which cannot be attributed to discourse effects. Theories which explain patterns of modifier-noun stress mostly or solely on the basis of FCA factors (such as that of Ladd 1984), are unlikely to be correct. We shall return to this issue at subsequent points in the paper. We devote the body of the paper to a survey of the "standard" pattern, as well as the systematic classes of exceptions to it.

2 The Standard Pattern

Before we take up the minority cases of left-dominant [*A N*] expressions and right-dominant [*N N*] expressions, it will be helpful to survey the standard forms of these constructions. Nearly all of the material discussed in this section is familiar from the literature, but since the full range of issues is rarely found discussed in one place it seems useful to review the facts here.

[4] We require that both the head noun and its modifier be content words, thus disallowing anaphors and excluding a few easily predictable [*A N*] FCA examples involving head words like *one*.

[5] Maidment (1989) has noted that news reporters often tend to place righthand stress on [*N N*] sequences that he feels should properly be left-stressed, and speculates (p. 187) that "this feature of broadcast speech is due to a desire ... on the part of the broadcaster to defer the intonational 'payoff' for as long as possible in order to create suspense or to make the news item sound more portentous." Indeed, Bolinger (1972, p. 643) notes that similar effects can even interfere with normal lexical stress for non-compound words. This possibility adds another possible dimension to FCA effects on which we will not comment further.

2.1 Standard Compound Nouns: N^0 [N N] Expressions

It is a well-known fact that English permits the free formation of compound nouns, the commonest type being of the form [N N]. The orthographic conventions for encoding compounding are varied. Short, frequent or fossilized compounds are often written as single words:

(4) drugstore, icewater, postman, gamecock, basketball, bathroom, ashcan, bartender, poolhall, earwax, lawnmower, flagpole, marshmallow, wallpaper, keyboard, waveform

Sometimes (especially in attributive position) a hyphen is used, and sometimes the two words are written separately: *masthead*, *mast-head*, *mast head*. The correlation between typographical practice and the semantic regularity of the resulting compound is at best imperfect. Some compounds that are commonly written as single words have a meaning that seems fairly compositional (5a), while other cases that must be written as two words have a special meaning that surely must be lexically listed (5b):

(5) a. phonecall (a call on the phone)
b. overseas cap (a particular style of cap)

The frequency of the compound, and the length of its constituent words, seem to be more important factors than semantic compositionality.

To some extent, the typography is a matter of style. Some writers prefer words written solid, while others like hyphens, or find pleasure in spaces. However, the typographical usage in actual text is quite variable. Even individual writers are not always consistent—within a single chapter of a textbook on the automobile electrical system, we have found *spark plug*, *spark-plug*, and *sparkplug*. However they may be spelled, such compound nouns are generally pronounced with primary stress on the first element. Indeed when the compound is run together or spelled with a hyphen, first-element stress is almost inevitable.

Because noun compounding is an easy way to create terms of art, technical writing, in the broad sense, is especially prone to such coinages. However, ordinary life also provides plenty of examples—a few days of recording those that came up in reading and listening produced over 5,000 examples, a few of which are given below:

(6) drug abuse, line backer, tool cabinet, feast day, knife edge, crop failure, dart game, body hair, shoe imports, phone jack, tea kettle, heat lamp, utility man, node name, post office, soap pad, printer queue, boat race, cocktail sauce, folk tale, land use, property value, star wars, junk yard, combat zone

2.2 Semantic Relations in N^0 Compound Nouns

The productive types of N^0 [N N] compounds fall into a number of different categories. As an initial cut, we will distinguish two broad classes on the basis of the semantic relation that holds between the two nouns. In the first

type, a paraphrase of the meaning has the first noun providing an argument for a predicate associated with the second (head) noun; we will call this type ARGUMENT-PREDICATE compounds. A typical example would be *lion tamer*, paraphrased as 'one who tames lions'. Extensive work has been done within theoretical linguistics on these so-called SYNTHETIC COMPOUNDS; see for example, Roeper and Siegel 1978, Selkirk 1982, Lieber 1983, Sproat 1985, Levin and Rappaport 1992, Marantz 1989. All of this work treats the relation between the lefthand member and the head in argument-predicate compounds by analogy to the relation between verbs and their objects.

In the second type of compound, paraphrases of the meaning involve a predicate not implicit in either word, with a meaning like *POSSESSION*, *PURPOSE*, etc., for which the two elements of the compound provide arguments. We will call this type ARGUMENT-ARGUMENT compounds. Some examples are *pie chart*, paraphrased as 'a chart that is like a pie', and *keyhole saw*, paraphrased as 'a saw used to make keyholes'. The semantics of these has been much less extensively studied, but Lees 1960 and in particular Levi 1978 are two works which discuss these compounds at length.

2.3 Argument-Predicate Compounds

The head noun in the *argument-predicate* type of compound may be an "agentive" *-er* nominal, a gerund, a derived nominal, or a noun without a verbal counterpart whose meaning nevertheless seems to put it in this category. A generalization which holds over all of these cases is that when the lefthand member is assigned an internal thematic role—typically whatever argument is normally assigned to the direct object of the verb from which the compound's head is derived—the main stress is on the lefthand member of the compound. The compound is therefore an N^0 by our assumptions.

Compounds with "agentive" heads. By these we mean, of course, examples such as the following, where in each case the lefthand member is interpreted as the internal argument of the verb from which the righthand member is morphologically derived via affixation of *-er*:

(7) shock absorber, torch bearer, syntax checker, car dealer, grain exporter, fire extinguisher, lens grinder, door knocker, deer hunter, rocket launcher, steel maker, can opener, music publisher, paint remover, knife sharpener, opium taker, window washer

It is easy to find (or think of) hundreds of other natural and familiar-seeming examples. There are a number of cases where quite large sets of objects will go nicely with a particular head. In the cases in (8), X could quite plausibly be any one of thousands of things:

(8) a. X maker: X = anvil, arrow, battery, bobbin, buckle, button, car, carriage, cheese ...

b. X supplier: X = asbestos, cable, copper, leather, lumber, paper, pipe, steel ...
c. X collector: X = art, book, car, clock, coin, mirror, pottery, stamp ...

There are relatively few values of X for which it might not, under some circumstances, be appropriate to speak of an X *maker*, an X *supplier*, or an X *collector*.

All of the above agentive compounds have lefthand main stress since their lefthand members function as internal arguments to the head. Needless to say, any agentive nominal that can occur alone, can also occur in constructions in which a preceding noun does not function as an object for the agentive's underlying predicate, but instead has one of the other sorts of relation possible in noun phrases. These constructions may show main stress on either member, depending on their nature:

(9) a. Left-dominant cases: ghost writer (does not write ghosts), baseball writer (does not write baseballs)
b. Right-dominant cases: Virginia creeper (does not creep Virginia), gas drier (does not dry gas), girl swimmer (does not swim girls)

This is expected given that agentive nouns are (after all) nouns, and should therefore be able to function like non-agentive nouns no matter what other properties they may have. So, parallel to *girl swimmer* there are other right-dominant N^1 constructions with appositive interpretations, such as *boy athlete*. Examples such as those in (9) cause problems for computational analysis, since it is often hard to be sure that an argument-predicate interpretation is wrong.

Agentive-headed argument-predicate compounds are not ordinarily well-formed if the left member is a measure noun or other pseudo-object. Nor is there normally any way to incorporate arguments whose expression normally requires a preposition or particle (though one finds a few examples like *city-dweller*, *church-goer* and *looker-upper*):

(10) *This meeting looks like a day-laster.
*Smoking-stoppers tend to be irritable.
*a water-looker-for
*a for-water-looker
*a water-looker

The lefthand member of agentive compounds may be modified in various ways, but it usually may not be quantified, usually may not be plural, and certainly may not have its own determiner or other noun phrase specifiers. Absence of determiners and other specifiers for modifiers is generally required in English, and this is a point we will take up again later. On its face, this suggests that the modifier for agentive nominals may be either N^0 or N^1, but no higher projection of N (and no projection of D) is allowed (see Fabb 1984):

(11) a toxic waste dumper
*a three dog owner
*a dogs owner
*a the Grand Canyon admirer

Compounds with gerunds as heads. English gerunds participate in a variety of constructions; in the one that concerns us here, a gerund is the head of a compound noun, whose lefthand member (usually) functions as the object:

(12) cattle breeding, ale brewing, carpet cleaning, number crunching, cost cutting, tape dubbing

Note that these cases are sometimes difficult in practical contexts to distinguish from the phrasal *NP V-ing* construction and it is certainly not difficult to construct genuinely ambiguous examples:

(13) The man eating shark was repulsive.

Sometimes an *-ing* nominal acquires an additional meaning that partly or entirely supersedes the act/process one; this meaning often denotes the result of the act or process—e.g., *dropping*, *building*, *writing*—or some materials or methods central to the process or action—*caulking*. Such cases do not usually form compounds of the type exemplified in (12), although (like all nouns) they participate freely in other sorts of compounding and modification. That is, a preceding noun will not serve as object if a gerund head is used in the "result" sense. Thus a *mouse dropping* is normally a dropping that comes from a mouse. If punk pilots adopted the practice of showering public gatherings with thousands of live mice, *mouse dropping* would be a natural way to refer to this unnatural act, but would not serve to denote its pitiful result, which we would have to call *dropped mice*.

As in the case of agentive-headed compounds, the argument-predicate compound constructions are typically ill-formed with pseudo-objects or with arguments expressed by means of prepositions:

(14) *He is capable of day-waiting.
*Cigarette-quitting is hard work.
*Oil-drilling-for is chancy.

Argument-predicate compounds with derived nominals as heads. As is well known (see Chomsky 1970, Thomason 1985, Sproat 1985, Safir 1987, among many others), derived nominals can have meanings that relate to an act or process associated with the related verb, or to the cause, instrument, method, resulting state of such an act or process. The second type of meaning—result, etc.—seems more erratic, while the first type—act, process or event—is more regularly found and its gloss is more easily predicted. We will follow such previous work by distinguishing these two types of meanings as PROCESS NOMINALS and RESULT NOMINALS.

In a compound noun headed by a process-nominal, the left member may correspond to a noun phrase in various syntactic relations with the related verb:

(15) a. Subject: cell division, commando raid, sunrise, police action, snake bite, proton decay, ether drift, moon glow
b. Object: nest construction, dream analysis, haircut, birth control, dress design, steel production, office management, heart massage, heat regulation, trash collection

The subject types often have (though clearly do not always have) right-dominant stress patterns:

(16) enemy invasion, police intervention, staff attempts, faculty decision, student inventions

Argument-predicate compounds without a deverbal head? There are many examples of $[N\,N]$ compounds whose head is not deverbal (and indeed may not have any corresponding verb) but seem nevertheless to correspond semantically to a predicate that takes other nominal constituents as arguments. Such words seem to form compounds of the argument-predicate type. Sometimes there is a corresponding expression with a postnominal PP expressing the argument, and sometimes not:

(17)	expert in ballistics	ballistics expert
	critic of music	music critic
	format of data	data format
	department of chemistry	chemistry department
	?broker of commodities	commodities broker
	*buff of opera	opera buff
	*thief of cars	car thief

Of course, as we shall see in Section 2.4, the range of argument-argument compounds is so broad that it is hard to be sure that examples such as those in (17) are not included in it.

Compounds whose left member is *self*. Compounds headed by agentives and derived nominals can freely occur with the word *self* as lefthand member. In all cases, the normal stress pattern is right dominant, for the same reasons—presumably FCA reasons—that reflexive pronouns are generally deaccented in phrasal contexts. Compounds headed by *V+ing* nominals do not so easily take *self* as a left member, although the corresponding compound adjectives are common:

(18) self-starter, self-promotion, self-igniting

2.4 Argument-Argument N^0 Compounds

We now turn to N^0 compounds where a paraphrase links the two words in the compound with a predicate not implicit in either one. We are limiting this category to endocentric compounds, so that their English paraphrase

will be something like 'an *N1 N2* is an N2 *relative-clause-containing-N1*, e.g., 'an *ankle bracelet* is a bracelet that is worn on the ankle,' or '*rubbing alcohol* is alcohol that is used for rubbing'. The range of predicates implied by such paraphrases is very large. Since this type of compound-formation can be used for new coinages, any particular compound will in principle be multiply ambiguous (or vague) among a set of possible predicates.

Consider *hair oil* versus *olive oil*. Ordinarily, *hair oil* is oil for use on hair, and *olive oil* is oil derived from olives. But if the world were a different way, *olive oil* might be a petroleum derivative used to shine olives for added consumer appeal, and *hair oil* might be a lubricant produced by recycling barbershop floor sweepings.

A coherent categorization is hard to find. In their everyday meanings, the compounds *olive oil* and *hair oil* resonate with many similar examples:

(19) a. Like *olive oil*

pattern *SOURCE-PLANT SUBSTANCE-THEREFROM-DERIVED*

peanut oil	sesame oil	safflower oil	soybean oil	corn oil
palm oil	cottonseed oil	corn syrup	bean paste	wheat bran
wheat flour	chickpea flour	rice flour	wheat bran	barley malt
corn starch	mango pulp	orange juice	lemon juice	apple juice
carnauba wax	guar gum	beet sugar	cane sugar	pine tar

b. Like *hair oil*

pattern *BODY-PART SUBSTANCE-THEREON-USED*

hand cream	skin cream	face powder	foot powder	lip gloss
eye drops	nail polish	underarm deodorant	hairspray	toothpaste

Such broad resonances have been the driving force behind classificatory schemes for argument-argument compounds of the type most fully developed in Levi 1978. A practical problem, discussed by Downing (1977), is that the set of patterns that would be required to achieve complete coverage appears to be open ended.

A more fundamental problem is that there does not seem to be any non-arbitrary way to decide on a single, coherent categorization of such patterns. The *olive oil* type of pattern exemplified in (19a),

SOURCE-PLANT SUBSTANCE-THEREFROM-DERIVED

covers substances such as oil, flour, syrup, bran, juice and so on, which are extracted by a variety of methods, use various portions of the source plant, and so on. This much generalization seems unproblematic—almost everyone would agree that these examples are instances of a type. However, the set can be extended in many directions. Each extension seems natural, but the resulting set is less and less coherent. One set of extensions (20a)–(20c) gradually relaxes the limitation of N1 to particular plants, allowing

for plants in general, to living things in general and finally to inanimate sources of materials; another (20d) allows N2 to be a count rather than a mass noun:

(20) a. *SOURCE-PLANT* is a more general category:
fruit juice, vegetable oil, grain alcohol
b. N1 is an animal:
whale oil, fish oil, chicken fat, horse hair, goose down
c. N1 is inanimate:
rock dust, river water, bread crumbs
d. N2 is a subpart that can be extracted or removed rather than a derived substance:
peach pit, chicken wing, rose petal, pine cone, coca leaf, corn cob, fish scale, corn husk, peanut shell

Plausibly, a general pattern to cover all the cases in (20) and the

SOURCE-PLANT SUBSTANCE-THEREFROM-DERIVED

instances in (19) would be

SOURCE SUBSTANCE-OR-PIECE-WHICH-COMES-FROM-SOURCE,

in which *COMES-FROM* includes derivation of a "new" substance by pressing, grinding, and so forth, of the source, and also by separation of a part from the whole.

Now consider the common cases where N1 is the whole of which N2 is a part:

(21) tire rim, mountain peak, arrow head, door knob, bed post, piano keys, shirt sleeve, table leg

These examples certainly are analogous to *peanut shell* or or *fish scale*, since both fit the rough schema **N2 is part of N1**. The examples like *olive oil*, in turn, are analogous to *peanut shell*, since both fit the pattern *N2 comes from N1*. But it seems less plausible that *olive oil* and *mountain peak* are in the same category. Each example, taken as a nucleus of generalization, yields several sets with which it shares some properties, but which may not share any properties with one another. A related example will give us new sets of neighbors, which overlap with the previous sets but are not exactly the same. Because of arguments like this, we doubt that an approach such as that proposed in Levi 1978 can be coherently and systematically pursued.

The 'connected-with' theory. As several writers have noted—e.g.; Dowty (1979)—the facts are consistent with a linguistic rule of argument-argument compounding that contributes only a vague 'connected-with' predicate, the more specific meanings arising from lexicalization and from the usual contextual circumscription of linguistically vague expressions. On this view, there is no well-defined hierarchical categorization of such

examples, since a given form may have analogical connections in many directions.

Thus the core meaning of *olive oil* would be something like 'oil connected with olives', and *hair oil* would be 'oil connected with hair'. In the case of these particular expressions, it is likely that their more specific meanings should be lexically listed. Novel cases—e.g., *thistle oil* or *moustache oil*—cannot get more specific meanings from a direct dictionary entry, but they can get a predisposition to go in one direction or another from analogy with the patterns in (19), the meanings of their component words, and common sense.

Words and fixed phrases easily acquire special meanings; utterances are always interpreted in context; and analogy with fixed expressions is a powerful determinant of everyday phraseology. So the mechanisms required by the connected-with theory are in any case available. By the nature of this theory, it cannot be disproved by positive examples, since (by some argument or another) it licenses any compound in any meaning. To show that this theory is wrong, we must show that some meaning relations are systematically excluded, and that some alternative hypothesis will distinguish those that occur from those that don't. This is nearly impossible to do without a precise account of such meaning relations,[6] which no one can at present provide.

A survey of argument-argument compounds. Our task is to distinguish the [*N N*] constructions with lefthand stress (N^0) from those with righthand (N^1) stress. Since the part-of-speech categories give us no help, any successful algorithm must rely in part on the semantic relationship between the words. The argument-predicate compounds previously considered were nearly all left-dominant. In this section, we will survey some common types of argument-argument compounds, limiting consideration to examples that are also left-dominant. After right-dominant expressions of form [*N N*] have been treated in a later section, we will discuss possible sources of the distinction.

Since any [*N N*] form could in principle have many meanings, we have chosen examples that rely on the existence of a meaning that is frozen through common use, that is implied by common sense given the word meanings, or that is strongly preferred due to the existence of a "schema" emerging from the analogical force of many related forms. It is an open question if such schemata achieve the technical status of linguistic "constructions" or formation rules, whether of morphology, syntax, or semantics. In practice, new instances of such compounds may usually be coined freely, whether by analogy or by rule. We divide the cases into categories that are designed mainly for expository convenience, and that are definitely far from complete. Fudge 1984 contains discussion of some of these classes.

[6]Though see Downing 1977 for some discussion.

(22) pattern *ELEMENT-IN-CAUSAL-CHAIN RESULT*:
drug deaths, heat rash, job tension, snow blindness, food poisoning, hay fever, tire track

For some reason, the head nouns in the pattern in (22) are usually negative in connotation. Persuasive examples with positive heads are hard to construct, and are generally right dominant, as in the example *for your listening pleasure*.

(23) pattern *TYPICAL-RESULT-OR-PRODUCT ELEMENT-IN-CAUSAL-CHAIN*:
disease germ, polio virus, growth hormone, honey bee, silkworm, song bird, oil well, sob story

(24) pattern *SUBTYPE TYPE*:
   a. N1 a name, N2 virtually redundant (hence the construction is almost exocentric):
   pine tree, cactus plant, collie dog, lilac bush, ivy vine
   b. N1 specifies species or type:
   preacher man, tree shrew, girlfriend, man servant, sports activities, tape measure, soldier ant

Note that *polio virus* is like *pine tree*, in that N2 is a generic biological classifier and N1 is a name for a particular type; however, a polio virus causes polio but is not a polio, while a pine tree does not cause pine, but is a pine.

(25) pattern *N1 THING-MADE-OUT-OF-N1*:
daisy chain, cable network, mountain range, grape arbor, chocolate bar, snowball, sugar cube (but cf. (68))

(26) pattern *POWER-SOURCE MECHANISM*:
steam engine, water wheel, vacuum cleaner, air brake, cable car (but cf. (75))

(27) pattern *INSTRUMENT ACTION* or *INSTRUMENT RESULT*
head butt, sword thrust, tank attack, knife wound, pot roast, pan fries, shovel cakes (but cf. (75))

(28) a. pattern *PLACE-WHERE-N2-IS-FOUND N2*:
   field mouse, mountain lion, desert rat, sea cruise, surface tension, farm boy, marsh gas, field trial, pond scum, house fly, street cop, city folk, alley cat
   b. pattern *TIME-WHEN-N2-OCCURS N2*:
   morning sickness, night blindness, Easter bunny, birthday party, Christmas present, morning coat, Sunday school (but cf. (70))

(29) pattern *N1 SOMETHING-WITH-PURPOSE-ASSOCIATED-WITH-N1*:
horse doctor, nose drops, fly paper, test pattern, cooking utensils, arms budget, plant food, face towel
(This is a large and diverse set that could be further subdivided.)

(30) a. pattern *THING-CONTAINED CONTAINER*:
picture book, photo album, gunboat, bear country
b. pattern *CONTENTS CONTAINER*:
parts bin, gin bottle, butter dish, olive jar, milk can, juice glass, garbage can, water bucket, stock drawer, linen closet, coffee mug, laundry basket, punch bowl, water tumbler, oil tank
c. pattern *N1 SHELTER-OR-TRANSPORTATION-FOR-N1*:
dog kennel, horse barn, cow shed, guest room, hay loft, corn crib, grain elevator, hay wagon

(31) pattern *SUBJECT-MATTER THING-FOR-WHICH-SUBJECT-MATTER-IS-RELEVANT*: tax law, abortion vote, budget debate, adventure story, love song, detective novel, oil crisis

(32) pattern *N1 THING-LIKE-N1*:
catfish, dragonfly, hermit crab, spider monkey, garter snake, kettledrum, frogman, cat burglar

Types and subtypes could be multiplied. For the reasons that we gave earlier, the categories are not disjoint and do not have sharp boundaries.

2.5 Compound Nouns whose Heads are not Nouns

English can also make compound nouns from various combinations of verbs, prepositions and nouns; see, for example, Selkirk 1982. Almost always, the stress pattern of these constructions is left-dominant. Also, most are written with a hyphen or as a single typographical word. These cases mostly fall outside the topic of this paper, but we list them briefly for completeness.

Phrasal verbs (verbs that combine with a "particle" or intransitive preposition) are nominalized freely. Common examples often have an idiosyncratic meaning:

(33) fall-away, run-away, throw-away, tear-away, slow-down, melt-down, come-down, splash-down, walk-out

Prepositions can also combine with verbs in the opposite order, although this type of combination is less productive:

(34) downdraft, downfall, downpour, downturn, downtrend, outburst, outbreak, outcast, outlook, outreach

Erratically, a verb may combine with a noun to form a noun. One [*V N*] pattern produces an exocentric compound whose referent is the (unexpressed) subject. This type is common in Romance languages but not in Germanic ones. It is no longer very productive in English, although it seems to have been popular in earlier times:[7]

[7]Marchand (1969, pp. 37–39) makes the interesting observation that such constructions are almost invariably pejorative, which seems to be true of the examples in (35) with the exception of *dreadnought*, which might explain its tendency to be less well-known.

(35) cut-throat, pick-pocket, pinch-penny, do-nothing, kill-joy, know-nothing, dreadnought, sawbones

A second $[V\ N]$ compound type is endocentric (that is, the referent is identified with the head noun). The semantic relation between the verb and the head noun is quite varied. Furthermore, it is often difficult to be sure that the lefthand member is indeed a verb and not the homophonous noun. New examples of this type of compound are frequently coined, but not all attempts are plausible. Thus a punch intended to stun might be called a *stun punch*, but a remark intended to offend could not be called an **offend remark*:

(36) searchlight, dancehall, springboard, workbench, grindstone, plaything, push pin, stuff bag, dive plane, trim tab, thrust plate, stab wound, snap bean, pitchfork, stun gun, hit man, push rod, lock washer

2.6 Compounds with Complex Parts, and the Compound Stress Rule

We have discussed binary nominals which are syntactically N^0 and which are therefore assigned lefthand stress. We now wish to consider the stress properties of N^0 constructions which have more than two leaf nodes. It is common to find compounds made up of subconstituents with internal structure. When the first member is itself a noun compound, and the second member is lexically simple, we generally find stress on the left:

(37) Air Force Academy, football game, money-market account, post-office box, data acquisition board, flashlight battery, X-ray film, bit vector machine, fund-raising operations

In each of the examples in (37), a compound noun is combined with another noun on its right, in a second level of compounding. In the result, the main stress is typically on the leftmost element. Sometimes, we find another recursion on the same pattern, in which a doubly compound lefthand member is further compounded with a single word after it. These are mercifully rare:

(38) water supply network repairs, windshield wiper blade replacement, error correction code logic

It is also possible to combine a noun on the left with a compound noun on the right, in a right-branching structure. These are perhaps slightly less common than the comparable left-branching structures shown in (37).[8] They often occur with a compound lefthand member as well:

(39) radio direction finder, spark-plug heat range, sink spray head, VAX instruction set

[8]For Swedish, Blåberg (1988, p. 68) claims that the right branching structure is significantly less common than the left branching structure in ternary nominals.

In such constructions, the primary stress is typically found on the penultimate member. In general, for a noun compound [*N1 N2*], if N2 is not itself a compound, then it will give up its claim to main stress in favor of N1. However, if N2 is a compound, then it will retain main stress within itself. This rule was called the COMPOUND STRESS RULE (CSR) in Chomsky and Halle 1968; we shall also use the traditional term NUCLEAR STRESS RULE (NSR) to refer to "phrasal" or righthand stress.

Liberman and Prince (1977) suggested that this rule was the same one that assigns main stress in nouns in general. As they observed, the righthand element of a noun is stress-dominant if and only if it branches, where the nodes of the relevant tree are words in the case of compound nouns, and syllables or feet in the case of simple nouns. Hayes (1980) suggested that this regularity should be recast as a simple final-stress rule, with the rightmost unit ignored (extrametrical) in the case of nouns. Again, one (simple) word is ignored in the case of compounds, and one syllable in the case of simple nouns. Under any formulation, the rule may be applied recursively to generate patterns of stress. (The effects of the so-called RHYTHM RULE, which shifts some non-final stress relations in order to create a more regularly alternating pattern, must also be taken into account.)

Across many types of noun sequences, the predictions of the CSR are fairly well verified. In (40) we give a set of typical examples containing four or more nouns, with the contextually implied grouping indicated by parentheses. The reader is invited to consider the predictions of the CSR, and compare them to his or her own intuitions about how the examples should be read:

(40) a. [[starter [drive gear]] clearance]
b. [[[gear selector] [control rod]] adjustment]
c. [[[power generating] station] [[control room] complex]]
d. [government [[tobacco [price support]] program]]

We note in passing that the CSR's determination of main stress position underdetermines the phrase structure in such cases—for instance, the pairs of structures shown below yield the same main stress location (marked in boldface) in each case:

(41) [A [[**B** C] D]] [[A [**B** C]] D]
[A [B [**C** D]]] [[A B] [**C** D]]

Due to the semantically protean nature of the compound construction, it is also not always clear what the structure of complex compounds is. Thus *tobacco price supports* might be *price supports* for *tobacco*, or (almost equivalently) *supports* for *tobacco prices*. Putting the main accent on *price* (which seems to be most people's preference) settles the matter in favor of the first option. The example *government tobacco price support program*, however, might be either [*government* [[*tobacco* [*price support*] *program*]]] or [[*government* [*tobacco* [*price support*]] *program*]], without any effect on

the main stress location. Of course, the situation is usually much clearer. In the context of use from which we took it, the parsing of *starter drive gear clearance* is unambiguous.

Having now considered some of the types of N^0 [N N] constructions in English, and further elaborating on the stress facts for such constructions which have more than two leaf nodes, we return to the examination of binary modifier-noun constructions.

2.7 Adjective-Noun Expressions

The canonical sequence [A N] differs from [N N] in that its parent is N^1 rather than N^0. Usually an [A N] sequence takes main stress on the right, in contrast to the left-stressed pattern of the [N N] sequences in (6):

(42) clear advantage, miserable bastard, comfortable chair, small difficulty, bitter end, special feature, ample girth, heavy heart, complete idiot, last judgment, fluorescent lamp, slender margin, arrant nonsense, smooth operation, wonderful person, stupid question, hasty retreat, great skill, worthless trash, enriched uranium, macroscopic Volvox, useful widgets, rare yak, melodious zither

2.8 An Explanation for the Basic Pattern

The High Road of Phrasal Stress is: CSR in [N N], NSR in [A N]. Syntactic category is a very good predictor of stress patterns for [A N], and a fairly good predictor for [N N]. Given this observation, the simplest account would appear to be one which derives stress placement from some syntactic property of the construction. The account we have been assuming wherein CSR applies in N^0 constructions and NSR in N^1 constructions, is just such an account. To be sure, there are complications with this story. One complication is that we must assume that exceptions to the general pattern, to which we turn in the next section, are exceptional from a syntactic rather than purely from a prosodic point of view. However, as we shall see, the alternatives to this assumption seem even less attractive.

3 Exceptions to the "Standard Pattern"

We will present a catalog of exceptions to the High Road to Phrasal Stress, with some commentary on its meaning.

3.1 Left-dominant [A N] Sequences

Lexicalized [A N] sequences. Some of the left-stressed [A N] cases are apparently lexically re-analyzed as nouns—i.e., as N^0. Some of these are written as a single word, or separated by a hyphen rather than a space; as with [N N] compounds, the typographical conventions are variable. Some of these cases are metonymic epithets[9]—*hardhat*, *red neck*—but many are not:

[9]Often known as *bahuvrihi compounds*, after the traditional Sanskrit terminology.

(43) square knot, soft spot, sticky bun, wild man, tonic water, pink slip, red line, black belt, blue book, golden boy, cold chisel, blind spot, sweet spot, dry dock, straight edge, whitewash, clean room

Examples like those in (44) below, which take the usual NSR pattern (unless contrast or whatever intervenes), show that the CSR does not automatically apply to fixed expressions of the form [*A N*], even if their meanings are partly or entirely idiosyncratic:

(44) white lead, red oak, green onions, brown sugar, Irish stew, hard liquor, red herring, brown betty, blue moon, black death, musical chairs, crazy eights, heavy metal, old maid

In order to attribute the stress pattern difference to the parent category, we must accept that the expressions in (43) are complex words of category N^0, while those in (44) are fixed expressions or semantically idiosyncratic phrases of category N^1. The alternative, however, is to add some feature like [±CSR] to the lexical entries of phrases.

For a few phrases of this type—indeed, for nominals of almost any type—some speakers have lefthand stress while others have righthand stress or can say the phrase either way. This demonstrates that any description must make it possible to effectively specify the stress pattern of such phrases without changing their meaning very much. However, there are a number of subregularities suggesting that the stress pattern in these examples should normally be determined from the parent category, so that the left-dominant expressions will be of category N^0, while the right-dominant ones will (mostly) be of category N^1. In particular, constructions for which it is reasonable to assume an N^0 analysis in the general case also tend to be pronounced with the stress predicted by the CSR. For example, [*A N*] sequences that are adopted as epithets (vocative or otherwise) tend to acquire compound-type stress, even if the same phrase can easily also be used as a normal phrase with righthand stress:

(45)

Epithet context	Non Epithet context
Hey, **big** mouth!	You've got a big **mouth**.
Hey, **fat** head!	You've got a fat **head**.
No more Mr. **Nice** Guy!	You're really a nice **guy**

A reasonable analysis of such cases, it seems to us, is that epithet-formation involves turning a phrase into a word—an N^0—and that it is this fact alone which accounts for the typical accentual pattern of such phrases.

Brand names, place names and plant or animal names with [*A N*] structure also have quite a strong tendency to develop lefthand stress. This tendency is especially pronounced for metonymic names, but many others show it as well:

(46) a. Long Beach, Hungry Horse, Red Wing, Old Town

b. Big Boy, Redbook, Green Stamps, Pure Aire, Smooth Pore, Kleen Flex, Ful Flo, Final Air, Dri-Pak, Hi-Flo, E Z Kleen, Tiny C, Soft Cote, Common Lisp, Rite Aid
c. yellowtail, bluefin, bluebird, blackbird, redwood, yellowjacket, bluebell, bluefish, redfish

Again, it is reasonable to assume that in the general case, such terms have become lexicalized and re-analyzed as N^0, and that this fact accounts for their accentual patterns.

Right-dominance is normal, however, in plant and animal names that are semi-compositional; that is, where the righthand element is a class name such as *oak* and the lefthand element is a descriptive adjective such as *red*. By our analysis, such cases would have to be analyzed as N^1. This pattern generally requires a head of some specificity, such as *leek* or *eagle*—more general heads such as *fish* or *bird* are more likely to be found in the class (46):

(47) red oak, curly dock, wild leek, bushy aster, common flax, bald eagle, American coot, spotted rail

Some place names seem also to be semi-compositional, in this sense:

(48) Big Bend, American Falls

Finally, although most [*A N*] brand names are left-dominant, as exemplified in (46), a few are right-dominant. Many of these have polysyllabic righthand members; a few may be considered semi-compositional:

(49) Orange Julius, Best Western, Dutch Masters, Big Mac, Leading Edge, Silver Cloud

We suggest that the nominals that we are calling "semi-compositional" retain their N^1 structure, despite becoming a fixed expression with a special meaning.

What is lexicalization? Note that not all lexicalized two-word nominals wind up with left-word main stress. In connection with the issue under discussion, there are in principle three types of lexicalization that a phrase can undergo.

First, a phrase may acquire a special meaning or association that requires it to be listed in a phrasal lexicon of some sort (see Di Sciullo and Williams 1987). Thus the compositional meaning of some topical phrase, say *mutually assured destruction* or *album-oriented rock*, is not enough to pin down the extra associations or technical meanings that it has acquired. Under the present analysis, however, the fact that these phrases are normally given phrasal stress means that they must be (minimally) N^1.

Second, a phrase may be treated as dominated by a lexical category, although presumably retaining its internal structure, or at least its divisions into words.

Finally, a former phrase may loose some or all of its internal structure as well as becoming a lexical category. There is no question that this long ago happened to *Worcester* on its way to being pronounced as if it were written <wuster>. A more current example is *high school*. One of the authors speaks a dialect that raises and fronts the diphthong [ai] when it precedes a voiceless consonant within the same word, so that the vowel in *fife* is quite different from the vowel in *five*. This raising applies obligatorily to the first vowel in *high school*, but not typically to the first vowels of, say, *fly paper* or *pie plate*, suggesting that the common word *high school* has lost its internal word boundary.

There is an interesting twist to this account which we merely mention here. As we have noted, the normal rules of English word stress, for nouns without internal word-level structure, say roughly to put the main word stress on the last non-final heavy syllable. Thus for an (originally) two-word phrase whose second word has at least one post-tonic syllable, the stress pattern predicted for the third stage of lexicalization, in which the internal word boundary is lost, is in effect, main stress on the second word. So if, for example, *red herring* were a noun without internal word boundaries, its main stress location would be the same as if it were a noun phrase. The preponderance of monosyllabic or final-stressed head words in (46), and of trochaic head words in (49), suggests that these lexicalized [*A N*] sequences may generally lack internal word-level structure, at least for purposes of stress assignment. That is, one can readily explain the distribution of stress patterns most simply by considering such cases to belong to the third stage of lexicalization outlined above.

Adjectives as substantives in [*A N*] "compounds." In other cases, left-dominant [*A N*] sequences seem perhaps to involve an adjective used as a substantive; see Levi 1978 for further discussion:

(50) athletic facilities, circulatory system, regulatory agency, electrical tape, erogenous zone, legal work, medical department, primary school, professional organization, solar system, tidal zone

All these are plausibly instances of productive patterns, in which a variety of other adjectives and nouns can occur in the lefthand position, with similar meaning. In some cases—e.g., *professional*—the adjectival form can also be used as a substantive, but often there is no common independently-usable substantive—e.g. *legal*, or the independent substantive has a different meaning—e.g., *primary*. Thus we cannot claim that the lefthand member of the examples in (50) is a true, independent substantive—a noun that happens to look like an adjective. However, the semantic connection between the two words in these examples seems similar to the relations that apply in analogous [*N N*] examples with left-dominant stress. So, the phrase *medical building* is essentially parallel to *chemistry building* or *office building*, except that 'medical stuff' has to be understood for *medical*.

Presumably the word *medicine* is avoided because of the confusion with the kind of medicine that comes in bottles.

The inadequacy of "implicit contrast": further arguments against FCA-only theories. We might try to explain left-stressed examples like those in (50) in terms of an "implicit contrast" between the modifier and its alternatives. This idea has no predictive value: there is no greater amount of "implicit contrast" in **legal** *work* than there is in *left* **lane**. To use the phrase *left lane* almost surely implies the alternatives *right* and *center*, and yet the main stress still falls on the second word. Here are some other examples of normally right-dominant [*A N*] sequences where the adjective is in "implicit contrast" with a small set of known alternatives:

(51) silver medal, heavy crude, young man, big business, large family, fresh bread

Without a definition of "implicit contrast" that can distinguish between the examples in (50) and those in (51), such explanations must be rejected as vacuous.

In some of the examples in (50) the head noun is a relatively empty one, so that one might appeal to some sort of redundancy-triggered deaccentuation. However, other [*A N*] expressions, with head nouns that seem equally empty, work more normally (i.e., by the NSR). In (52) we have tried to use heads that are both general and common, yet in all of these examples the primary stress falls on the righthand member:

(52) an enormous area, a nice person, a special place, a funny thing, a strange man, a loathsome object, the thirteenth time

Again, without a definition of "redundant head" that can distinguish (50) from (52), this account must also be considered to have no predictive value.

Conclusions about the left-dominant [*A N*] cases. So, we are left with two fairly clear patterns and some unanswered questions about them. There is clearly a pattern of adjectives-masquerading-as-substantives forming [*N N*]-like compounds with nominal heads, as in (50). There also seems to be a pattern of [*A N*] sequences becoming single words, and acquiring the stress patterns thus predicted, as shown by the examples in (46) and (45), and presumably the examples in (43).

What determines the set of adjectives that can be used as quasi-substantives? It is surely not a random choice. Why is the substantival usage often (not always) restricted to attributive position? Why do the examples with monosyllabic heads in (44), such as *Big Ben*, *blue moon*, *hard times*, fail to "collapse" into the single noun stress pattern along with *Big Foot*, *blue bird*, *hard ball*, and all the others? In order to make progress with these questions we probably need a substantial fragment of the relevant "phrasal lexicon," giving us a large enough list of examples to be able to see patterns and test predictions.

We have motivated a distinction between "semantic" lexicalization (that is, a fixed expression with a non-compositional meaning) and "syntactic" lexicalization (that is, a fixed expression that is stressed as if its parent category was N^0 rather than N^1). It appears that the second implies the first, but not vice versa. Finally, we have distinguished the category of "morphophonemic lexicalization," in which a lexicalized compound word loses its internal word structure, and becomes subject to phonological and phonetic processes that would ordinarily only apply word-internally. We assume that this can only happen to expressions that are already syntactically lexicalized. We have also suggested that many of the lexicalized [*A N*] sequences, discounting the cases discussed above where the adjective is functioning as a noun, may fall into this third category, though it may be difficult in particular cases to distinguish between this possibility, and the possibility that the case in question really belongs to the semantic or syntactic lexicalization class.

3.2 Some Other Left-dominant Phrasal Units

Besides the atypical [*A N*] units just discussed, there are a few other cases of non-[*N N*] noun phrase structures that are left-dominant. We survey them below.

Compound-like cases with possessives as left member. Sometimes a phrase of the form [*N's N*] or [[*N N's N*] regularly prefers stress on the first member, although such phrases normally are stressed by the NSR:

(53) cashier's check, bachelor's degree, servant's entrance, pig's feet, ram's horn, chef's knife, lubber's line, user's manual, goat's milk, bosun's chair, machinists' union, sheep's wool, auto workers' union, meat cutters' union, bull's eye, crow's nest

A familiar set of accounts are available here: lexicalization; implicit contrast of the left constituent; low information content for the right constituent. As in the case of left-dominant [*A N*]] expressions, all of these accounts have predictive problems.

Thus, although many of the expressions in (53) are semantically lexicalized, there are other fixed or idiomatic [*N's N*] expressions that are nevertheless right dominant:

(54) pope's nose, cat's pajamas, fool's errand, busman's holiday, blind man's buff, Zipf's Law

So, as outlined for the [*A N*] examples discussed above, lexicalization in and of itself does not predict N^0 status.

Still, there is some evidence that lexicalization has a part in such examples: again [*N's N*] epithets are often left-dominant (though one of the authors has righthand stress in *horse's ass*) as are brand names, idiomatic plant names, etc.:

(55) a. momma's boy, cat's paw, dog's body

b. Boar's Head, Bird's Eye, Woman's Day
c. Bachelor's-button, Baby's-breath, Lady's-slipper, Solomon's-seal, Shepherd's-purse, Dutchman's-pipe, Prince's-feather, Tinker's-weed, Lamb's-quarters, Goat's-rue, Lion's-foot, Virgin's-bower

It is interesting that metaphorically descriptive names, such as those in (55c), are inevitably left dominant, while a right-dominant treatment is given to species names in which the discoverer's name possessively particularizes the name of the larger class. These cases are identical in spirit to the "semi-compositional" [*A N*] names surveyed in (47):

(56) Swainson's thrush, Lincoln's sparrow, Kirtland's warbler, Catesby's Trillium

The parent node of left-dominant possessives. The possessives in left-dominant [*N's N*] expressions have the distributional properties of position 1 modifiers: certainly, at least, they can be preceded by articles, quantifiers, adjectives, and nominal modifiers. This is in striking contrast to the evidence supporting the view that possessive phrases normally occupy the same peripheral determiner position as articles and demonstratives—what we called position 4 in (1):

(57) a. the fresh mare's milk
b. 16 Unix user's manuals
c. a physics master's degree
d.*the red Sam's car
e.*16 well-thumbed Knuth's books.

Since *Unix manual* and *physics degree* are ordinarily left-dominant, examples like (57b) and (57c) suggest position 1 as opposed to position 2 status for *Unix* and *physics*, respectively, and from this we conclude that *user's* and *master's* must also occupy position 1.

The possessive phrases in (56) pattern more like position 2 elements than like normal possessives, occurring inside position 3 and other position 4 elements:

(58) a. the three Catesby's Trilliums
b. yesterday's innumerable Swainson's thrushes

We have not found any convincing examples to suggest that these right-dominant possessives, as opposed to the left-dominant ones, can occur inside position 1 modifiers. Some of the other right-dominant lexicalized [*N's N*] expression are similarly distributed:

(59) the coveted pope's nose, an obvious fool's errand

These striking distributional differences make it reasonable to treat lexicalized possessive phrases as having a different parent category from normal possessive phrases. The most plausible candidates seem to be N^0 for the left-stressed cases and N^1 for the right-stressed cases.

Implicit contrast in left-dominant possessives? In many of the expressions in (53), the possessive is in some sense implicitly contrastive; thus *lamb's wool* may be distinguished from the wool of sheep, goats, or whatever. However, the same thing may be said of nearly any [*N's N*] construction. In the film title *The President's Analyst*, the expression might be said to distinguish the President's analyst from the analyst associated with anyone else, but the expression is stressed on the final word. The same can be said for any of the right-dominant possessive phrases in (54) or (56). Thus the "implicit contrast" treatment of these expressions has no predictive value, nor any apparent descriptive appropriateness, and should be rejected as a general explanation.

Left-dominant [*N PP*] expressions. There are a fair number of compound-like *N PP* words that take main stress on the lefthand element:

(60) jack-in-the-box, brother-in-law, mother-in-law, free-for-all, hole-in-the-wall, good-for-nothing, stick-in-the-mud, snake-in-the-grass

These cases are probably related to the general tendency of NP-locative sequences to deaccent the locative phrases, regardless of whether or not the phrase is idiomatic, and of whether the locative is a modifier of the preceding *NP* or an adjunct of an earlier verb:

(61) a. Here's **mud** in your eye.
b. He's got a **bee** in his bonnet.
c. He's got a **chip** on his shoulder.
d. I've got a **spot** on my shirt.
e. There's a multi-legged **creature** on your shoulder.

However, many fixed or idiomatic expressions of the form *N PP* (or perhaps *NP PP*) take righthand stress:

(62) man of the world, bolt from the blue, chip off the old block, drop in the bucket, grant-in-aid, jack of all trades, king of the hill, man in the moon, man o' war, manna from heaven, pie in the sky, piece of cake, tug of war, word of mouth

Again, epithets seem disposed to be left-dominant:

(63) Little Miss Nose-in-the-air, Mr. Full-of-fun, light-of-my-life

Other cases. Occasionally a verb phrase or similar collocation takes on an idiomatic nominal meaning, along with main stress on its leftmost word. These examples are probably related to the cases in (35):

(64) forget-me-not, thank-you-ma'am, know-it-all

However, most verb phrase nominalizations are right-dominant. Game names are a common source:

(65) spin the bottle, pin the tail on the donkey, squeak-piggy-squeak, go fish, beggar my neighbor, follow the leader, catch-me-if-you-can

Likewise most nominal uses of other phrasal constituents are right dominant:

(66) I spy, Simon says, odd man out

The kitchen sink: a potpourri of phrasal lefthand members. Here is as appropriate a place as any to note that in some informal styles, various phrasal categories can be freely used as prenominal modifiers, with an appropriately generic meaning. Verb and adjective phrases are particularly common. This usage permits free inclusion of pronouns, articles and other things that are usually forbidden in modifiers. Many such phrases—*top-of-the-line*, *hole-in-the-wall*, *turn-of-the-century*—are fixed expressions, but nonce formations do occur. Examples are extremely common in certain journalistic styles, from which the following examples are all taken.[10] Both left-dominant and right-dominant stress patterns occur in this group:

(67) an old-style white-shoe do-it-on-the-golf-course banker, the usual wait-until-next-year attitude, a wait-until-after-the-elections scenario, a kind of get-to-know-what's-going-on meeting place, the like-it-or-lump-it theory of public art, state-of-the-union address, a 24-hour-a-day job, a 1-percent-of-GNP guideline, a run-of-the-mill meeting, a sweep-it-under-the-rug amendment, a middle-of-the-road format, the state teacher-of-the-year title, a take-it-or-leave-it choice, the yet-to-be-written 1987 bill, a certain chip-on-the-shoulder attitude, make-it-from-scratch traditionalists, Speak-Mandarin-Not-Dialects Month, a rob-Peter-to-pay-Paul system, the nothing-left-to-chance approach, get-out-the-vote drives, the don't-trust-anybody-over-30 crowd, national clear-your-desk day

Such examples are quite consistent with our approach to predicting stress from structure in complex nominals, suggesting that both positions 1 and 2 may be occupied by elements projected from (nearly) any lexical category.

3.3 Right-dominant [*N N*] Expressions

In the texts mentioned earlier in this paper, about 25% of the [*N N*] phrases had main stress on the right, and essentially none of these had a plausible FCA account. We argue that these are instances of N^1.[11] In this section we outline some of the semantic relations which one finds in these cases[12]

Noun modifiers expressing composition. Some right-dominant [*N1 N2*] sequences are of a familiar type that means something like 'N2 made of N1', 'N2 made with N1', or 'N2 with N1 as a featured part':

(68) pattern *N1 THING-MADE-OUT-OF-N1*:
rubber boots, steel plate, duck soup, gold medal, corduroy suit, brass

[10]Lieber (1988, 1992) also discusses examples of this kind, arguing that one needs to allow phrasal entities inside compounds in the general case.

[11]Again this is a traditional position: see Bauer 1983, pp. 104–108 for a discussion.

[12]See Fudge 1984 for a discussion of some of these classes.

bed, diamond ring, nylon rope, plaster cast, wax figure, asbestos tile, meat pie, chicken gumbo, wood floor, mushroom omelet, carrot halvah, rice pudding, apricot jam, corn tortillas, squash pie, beef burrito

In some cases of common words, 'made with N1' [*N1 N2*] sequences revert to lefthand stress, presumably due to lexicalization of the same type that is involved in (43); one may refer back to (25) for some examples, or consider:

(69) banana bread, rye bread, peanut butter, corn chips, potato chips, butter cookies, oatmeal cookies, orange juice, coconut milk, aspirin tablet, chocolate bar, chocolate milk, cornstarch, silverware

As expected if lexicalization is at issue, there are many individual differences in this area; some people say *peanut bu'tter* in place of *pea'nut butter*, while others have *potato sa'lad* in place of *pota'to salad*, *chocolate mi'lk* in place of *cho'colate milk*, etc.

In (69), note the prevalence of left-dominance in the *cookie* and *bread*-head cases, even for new coinages like *zucchini bread* or *cinnamon cookies* (the latter for one of the authors only), and compare this pattern to the results of combining the same modifiers with the head *pie*. The left-dominance of bread terms even extends to some forms with adjective modifiers, such as *sticky bun* and *hard roll*. These cases, in which a particular head word has a more-or-less strong affinity for a particular stress pattern, are similar to the familiar case of *Street* versus *Avenue*, discussed below.

Time, place, class, etc. In other cases, N1 is a time or place, loosely defined:

(70) a. pattern *PLACE-WHERE-N2-IS-FOUND N2*:
garage door, basement walls, attic roof, kitchen sink, cell window, pantry shelf, library curtains, bedroom furniture, cell membrane, mountain pass, valley floor, college president, city employee

b. pattern *TIME-WHEN-N2-OCCURS N2*:
summer palace, fall weather, winter carnival, summer sausage, autumn leaves, spring flowers, winter cold, spring cleaning, September morning, January thaw, Christmas dinner

Proper name modifiers. Closely related are the cases in which a proper name is used as a modifier, to establish location, type, period, source, ownership, brand or whatever:

(71) pattern *PROPER-NAME THING*
Staten Island ferry, Connecticut Yankee, US ambassador, West Texas barbecue, Busch beer, Napoleon brandy, Pennsylvania crude, AT&T headquarters, India ink, Tiffany lamp, GB stronghold

Of course, full proper names go by the NSR, as do compound names for locations, other than those that end in *Street*, or *Land* such as those in (72c):

(72) a. George Washington, Barnacle Bill
b. Newark Airport, Chesapeake Bay, Bryce Canyon, Park Drive, Baxter Estates, Niagara Falls, Strawberry Fields, Union Gap, Yosemite Lodge, Tuolumne Meadows, Adygei Autonomous Oblast, Abiathar Peak, Cat Mousam Road, Times Square, Sturbridge Village, Blue Jay Way
c. Bank Street, Arnhem Land

Generally, names for institutions, organizations, and so forth, also go by the NSR, though there are exceptions (73b):

(73) a. Harvard College, Locksley Hall, Coolidge High, Bell Laboratories, Harvard Law, Widener Library, MIT Press, Bethlehem Steel, Stevens Tech, Yale University, Harvard Yard
b. Dudley House

Measures, methods, media. Another case of right-dominant [*N N*] phrases includes various sorts of measure-phrases in the N1 position:

(74) pattern *MEASURE THING-MEASURED*:
mile run, pound note, gallon jug, pint jar, two-alarm chili, three dog night, four-man front, two-minute warning, eight-hour day, six-figure salary, fixed-length record, 12-gauge shotgun, three-day pass, three-minute egg

One apparent exception to this pattern is **minute** *steak*. However, the difference between the bulk of the examples in (74) and *minute steak* is that that the former are genuinely compositional instances of measure-phrase modification. That is, the lefthand material is measuring the head on an appropriate (if sometimes idiosyncratic) scale. So, a *mile run* is a run which is really a mile long, a *three-minute egg* takes three minutes to cook, *two-alarm chili* is placing the chili in question at a reasonably high point on a scale of potential gastric disturbance, and a *three dog night* gives a measure of the ambient temperature of the night in question using the number of dogs required to keep oneself warm as a metric. The term *minute steak*, however, does not mean that the steak in question takes one minute to cook, but merely that it cooks quickly.

Also, we have right-dominant [*N1 N2*] constructions with characteristic methods, media or energy-sources in the N1 position, although the behavior of this class is erratic; see left-dominant examples in (26) and (27):

(75) pattern *METHOD-OR-MEDIA-FOR-N2 N2*:
bathtub gin, gas chromatograph, electron microscope, pinhole camera, gasoline engine, kerosene heater, propane torch, shotgun wedding, gunboat diplomacy, jet engine, steam radiator, microwave popcorn

Residual examples. Other examples of right-dominant [*N N*] phrases belong to patterns whose classification is not clear to us. We reproduce a list for illustrative purposes:

(76) fly ball, discount bookstore, dictionary definition, capital gains, weather helm, industry leader, color monitor, trial run, combination lock, sacrifice single, round-trip ticket, gospel truth, battleship grey, blood relative, tramp steamer, return ticket, precision tool, race suicide, eggshell china, fossil man, sex education, rogue elephant, touch football, ball bearing, fullback draw, championship series, party line, cash customer, child labor, chain reaction, touch typist, home run, string quartet, pony express, parcel post, police custody, polka-dot dress, box score, world bank, industry leader, crash course, crash landing, crack regiment, smash hit, snap judgment, bum rap, sponge rubber, college degree, shoestring tackle, bit part, barrier reef, toy gun

Some of these examples may follow the pattern *N2* IS-A-*N1*, which typically follow the NSR as discussed for (9b): *rogue elephant*, *fossil man*, *sponge rubber*, *toy gun*. Notice, incidentally, that *color monitor* is stressed one way if you mean 'monitor that shows color'—that is, the usual *Byte* magazine reading, and the other way if you mean 'device or circuit for monitoring colors'—say, a new improvement in TV set technology, also known as the *spectrum inspector*. In the second case, we have a clear argument-predicate compound.

Left-headed sequences. All of the examples of modified nominals (with the exception of *bahuvrihi* constructions) which we have discussed so far have had the nominal head on the right and the modifier(s) on the left, the normal ordering for English. However, there are some cases where it appears to be necessary to assume that the head of the construction is on the left and the modifier is on the right. We will assume that this is simply a marked option, though it is certainly true that adjectives in English may quite freely appear on the right of the noun they modify under appropriate conditions: *a man tall, dark and handsome* (see Bolinger 1967). In many of the examples we will see, the particular head in question is apparently marked to occur on the left. All of the leftheaded examples are stressed by the NSR, hence are N^1 under our analysis. For example, (77) below contains left-headed constructions where the righthand member is an identifying name or number:

(77) vitamin C, route 1, brand X, exit 14, peach Melba, steak diane, Cafe Beethoven, Club Med

So, while a *garage door* is a door, and a *company president* is a president, *vitamin C* is a vitamin, not a letter or a programming language, and *steak diane* is meat. The final examples in (77) are either borrowed from French, or at least modeled on French syntactic patterns, which

are more single-mindedly left-headed than English ones. Note that food terms in particular are commonly (and productively) left-headed, and Romance influence is particularly likely in this domain: *beef Wellington*, *chicken cacciatori*, *lamb vindaloo*, *eggs benedict*, *bananas Foster*. Some of these may have questionable headedness: is *lamb vindaloo* a kind of vindaloo made with lamb, or is it a kind of lamb cooked with a vindaloo sauce?

As noted above, a number of these cases involve heads which seem simply to be marked to occur on the left: so if a new vitamin were discovered, and assigned the letter 'X', the vitamin would have to be called *vitamin X*, not **X vitamin*. Similarly, *exit 14* could not be called **14 exit*.

Further examples of constructions of that lean towards left-headedness, many of which instantiate productive patterns, are given below:

(78) pattern *CLASSIFIER NAME*:
Bayou Goula, Cape May, Chancellor Adenauer, Citizen Kane, Comrade Andropov, Dr. Smith, Fort Eustis, Key Largo, King George, Lake Erie, Lake Huron, Loch Ness, Mount Hood, planet Earth, widow Brown

The *CLASSIFIER NAME* pattern can be augmented as a class by the introduction of new titles borrowed from foreign languages, such as *Generalissimo Franco*, but it cannot be freely used for all classifiers, even for some that seem like minor variants of cases that work fine:

(79) *Pond Walden, *Hill Blue, *bride Smith

3.4 A Possible Reason for N^1 [N N]s: Is the nominal modifier an adjective?

Leaving the left-headed cases now, we now return to a possible explanation for the right-dominance of certain right-headed [N N] compounds, namely the traditional notion that in such cases the lefthand noun is functioning as an adjective. For example, for the *N1 THING-MADE-OUT-OF-N1* cases given in (68), we might postulate a null-affix derivational rule that makes adjectives out of nouns, since the usage usually generalizes to predicative patterns:

(80) The boots are rubber.
The plate is steel.
The medal is gold.
The suit is corduroy.
The bed is brass.

When the meaning is more like 'made with X as a relevant part', or whatever, the attributive noun cannot so felicitously be used in predicate position with its head as subject:

(81) ?The soup is duck.

And in the majority of the right-dominant [*N N*] sequences we have surveyed, the predicative counterpart is totally implausible:

(82) *The ball is fly.
*The jail is county.
*The door is garage.
*The membrane is cell.

All in all, the adjectival analysis of the lefthand noun in N^1 [*N N*] constructions runs aground on the problem of limiting the distribution of the resulting adjectives. A more straightforward idea is simply to generate the nominal modifier in the same phrasal slot that an adjectival one would occupy; that is, in position 2 of (1), which we have identified as dominated by N^1. The stress and the meaning of the nominally modified phrase ought then to follow without further ado, if the rules of stress assignment and semantic interpretation are correctly arranged.

3.5 The Case of [[*A N*] *N*]

As is well known among aficionados of complex nominals, although occasionally ignored by other morphologists, one can combine an [*A N*] unit with a nominal head in a semantic relation typical of a simple [*N N*] compound with the same head. In such constructions, the main stress would typically fall on lefthand member, CSR style, as for most [*N N*] compounds. Since the lefthand member itself is of the form [*A N*], its internal stress pattern normally goes by the NSR, and the result is main stress on the middle word:

(83) toxic waste cleanup, collective bargaining agreement, balanced budget amendment, civil rights bill, used car business, financial planning consultant, due process clause, floating underflow trap, systolic array machine, fresh fish shop

Both the existence of such examples and their normal stress pattern follow straightforwardly from our proposals.

4 The Structure of Premodified Nominals: A Summing Up of the Theoretical Issues

We have examined a wide range of modified noun phrases. In this section, we will summarize the structural analyses we have suggested for the various types of nominals that we have discussed. In addition, we will have a few things to say about some other nominal constructions in English and other languages.

We will then extend our arguments against what we have called FCA-only theories of nominal compound stress, by examining the approach of Ladd 1984. We will also discuss the distinction between morphology and syntax: should some of constructions we have been discussing be viewed as the product of a component of the grammar—the morpholog-

ical component—that is crucially different from the component responsible for phrasal composition? Many writers have argued that this distinction should be made, Di Sciullo and Williams 1987 being perhaps the most vocal supporters of this view. We will evaluate some of their arguments in terms of the phenomena discussed in this paper, and suggest that the evidence for such a split is not strong.

4.1 A Summary of the Structure of the Modifying Phrase

We have seen examples of English premodified nominals with various sorts of supra-lexical constituents as their lefthand member. At least the patterns $[A\ N]$, and $[N\ PP]$ are fairly productive. This suggests that the lefthand constituent of premodified nominals is (maximally) an X^1, where X is commonly (though not necessarily) N^1:

(84) $[[_{N^1}\ A\ N]\ N]$ equal rights amendment
$[[_{N^1}\ N\ PP]\ N]$ balance of payments problems

Again, there are certainly restrictions. For example, N-plus-complement or A-plus-complement modifiers are somewhat restricted. There are stylistic issues, and common or lexicalized modifiers are preferred in these cases. Examples like these seem infelicitous:

(85) *Where's my [tree in the garden] clipper.
*John is a [proud of his son] man.

Examples like those in (85) are generally taken to be ruled out by a Head-Final Filter (Fabb 1984, Di Sciullo and Williams 1987), which states that the head of a modifier must be adjacent to the modified word. However, the fact that more common phrases seem unproblematic as modifiers (*balance of payments problems*) suggests that the Head-Final Filter may actually not be a grammatical constraint, but rather a parsing constraint: as long as the lefthand component is recognizable as a unit in the context of use, its exact construction is not as limited as constraints such as the Head-Final Filter would suggest.

As exemplified in (11), the left member of a noun compound can ordinarily not have its own independent article. The restriction against articles is pretty strong even for things that are clearly lexical items—thus we say *the Bronx Borough President*, not **the the Bronx Borough President*; see Fabb 1984 on these. Also excluded are noun phrases specified by a a *wh*-word, and any form of pronoun acting as the modifier on its own. Quantifiers, which are often taken to be specifiers of N^2 (see Abney 1987) are normally unacceptable. So, the following examples are all impossible:

(86) *an [every kitchen] table,*the [which dog] owner,*the [it] hater

However, note that modifiers containing numerals are not excluded:

(87) the four color theorem, the eight queens problem, three-mile mark, a $3 million offer

So, on the one hand we have clear evidence that in many cases the modifier must be phrasal, and we have equally clear evidence that the modifier cannot be a fully specified noun phrase. We can capture such restrictions quite neatly if we assume that prenominal nominal modifiers are maximally N^1. This will rule out modifiers such as *The Bronx*, *every kitchen*, *which dog*. To rule out pronouns such as *it* as modifiers, one could appeal to a suggestion of Paul Kiparsky (p.c.) that closed class items (at least in English) do not typically allow further morphological derivation. On this account, *it hater* would be bad for the same reason as a derived form such as *it-ishness* is bad. To be sure, other explanations are available: another line of thought says that projections of functional categories may not serve as modifiers. This rules out pronouns and also projections of D, such as *The Bronx*, as modifiers. Indeed, a similar set of restrictions holds of adjectival phrases: degree specifiers such as *so*, *that*, *how* are obligatorily absent from prenominal adjectival modifiers:

(88) a. The houses are that large.
*the that large houses
b. How large are those houses?
*Those how large houses are over there?
c. The houses are so large.
*the so large houses[13]

Abney (who discusses these facts) suggests, on analogy with his analysis of noun phrases, that full adjective phrases are really Degree Phrases (DegP) where the degree word is the head, just as a determiner is the head of DP, with the AP as a complement of the degree word; see Abney 1987, p. 321:

(89) $[_{DegP}\ [_{Deg}\ \text{so}\ ]\ [_{AP}\ [_{A}\ \text{nice}\ ]]]$

One can then rule out the examples in (88) by assuming that only APs—possibly, in fact, maximally A^1, can appear in prenominal position. This is essentially what Abney argues, although, his analysis makes the rather novel assumption that the adjective is the head in a construction like *big dog*, that this construction is therefore an AP, and that the NP *dog* is a complement of *big*. We have of course taken the more traditional view that canonical prenominal modifiers, namely the prenominal adjectives, are adjoined to N^1 in English.

So, it seems that for the canonical prenominal modifiers, namely projections of N and projections of A, there is some evidence that such modifiers are typically maximally single bar projections—X^1. If we look again at the examples of highly-complex modifiers in (67), we see that they are actu-

[13]One can say things like *how large a house*, *so large a house*. Such constructions are limited, in the sense that only an indefinite singular NP can follow the degree-marked adjective: **how large that house*, **how large houses*. Abney argues that these are adjectival phrases (DegPs) where the head adjective takes a fully specified noun phrase as a complement.

ally consistent with this idea—they may be projected from non-traditional categories such as V, but may plausibly be analyzed as being of level X^1 or below.

4.2 More on the Distinction Between N^0 and N^1 Modifiers

Having argued that the modifier of a noun is maximally X^1, we now overview the arguments for our assumptions about the structural relationship between the modifier and the head. It is worthwhile to start out by placing this assumption about modifier-head relations in a broader context of current ideas about phrase structure.

We have maintained that modifiers are adjoined to either N^0 or N^1. The diagram in (90) illustrates the structure we have assumed for an N^1 modifier such as the adjective *red*:

(90) $[_{DP}\ [_{D^0}$ the$]\ [_{N^2}\ [_{N^1}\ [_{A}$ red$]\ [_{N^1}\ [_{N^0}$ book$]]]]]$

We have assumed that most $[N\ N]$ compounds are instances of modifier-head constructions, in this case involving adjunction to N^0:

(91) $[_{N^0}\ [_{N^0}$ dog$]\ [_{N^0}$ house$]]$

However, we have argued at length that not *all* instances of $[N\ N]$ modification are at the N^0 level; similarly, not all instances of $[A\ N]$ modification are N^1. Thus we have posited structures like these:[14]

(92) a. $[_{N^1}\ [_{N^0}$ kitchen$]\ [_{N^1}\ [_{N^0}$ sink$]]]$
b. $[_{N^0}\ [_{A^0}$ blind$]\ [_{N^0}$ spot$]]$

Our primary arguments involved default stress patterns and some informal consideration of semantic relations, but there is also evidence from sequence constraints among modifiers. Note that one does not find $[N\ N]$ constructions which we would argue on stress grounds to be of the type N^1, occurring as the righthand member of a construction whose parent node ought to be an N^0. So, one can say **coffee** *jar* (N^0), and *plastic* **jar** (N^1), but one cannot get **coffee plastic* **jar** (cf. *plastic* **coffee** *jar*) since the structure would have to be $*[_{N^0}\ N^0\ N^1]$, which is ruled out by X-bar considerations given that the rightmost member of a compound in English is generally the head (see (105) below).

On the other hand, complex expressions which look on the surface to be phrasal constructions, but which we have argued on stress grounds to be N^0, can occur in the righthand position of compound nouns. The complex

[14]It is worth noting that the structure in (92a) is vaguely related to the proposal in Selkirk 1984 (pp. 43–50) that right stressed compounds involve adjuncts, where the modifier is in some sense more external than modifiers in a left stressed compound, which are taken to be arguments. Selkirk, however, also seems to suggest that at least some compound-like constructions, such as *steel* **warehouse** (i.e., warehouse made of steel) might be viewed as phrasal collocations (p. 247) as opposed to compounds of the form *ADJUNCT-HEAD*. It is not clear from her discussion how one is supposed to distinguish these cases.

examples in the third column are all right-branching N^0s having primary stress on the penultimate member as predicted by the CSR:

(93) **user's** manual **Unix** manual [Unix [**user's** manual]]
dry dock **trawler** dock [trawler [**dry** dock]]

Needless to say, prenominal noun modifiers that on our analysis are dominated by N^1, are generally reasonable candidates to stack outside adjectives that occur at the same level. Thus the right-dominant character of *plastic cup* is consistent with the possibility of *plastic disposable cup*, though the alternative ordering may be slightly preferred.

By giving a structural explanation for the unavailability of examples like **coffee plastic jar*, we are claiming that such examples are ill-formed for the same kind of structural reason that **difficult the problems* is ill-formed. It is worthwhile contrasting such rigid ordering restrictions with the much softer ordering restrictions on prenominal adjectival modifiers, which are all introduced at the same (N^1) level, but which nonetheless display well-known ordering preferences:

(94) large red ball ?red large ball
oversize square table ?square oversize table

One could, of course, ask whether these ordering restrictions should also be handled by assuming intermediate structural positions or levels? For example, measure adjectives like *large* could be placed in a slot "outside" color adjectives like *red*, since the former prefer to precede the latter. However, it has been observed in the literature on prenominal adjective ordering (see, most recently, Sproat and Shih 1990b) that these ordering restrictions are only preferences, operating in the absence of pragmatic reasons to choose a different order. It is perfectly possible to choose an alternative order, given appropriate context:

(95) I want the *red* large ball, not the *blue* one.

So it seems best to agree with the traditional view that these ordering preferences reflect some sort of natural continuum of psychological affinity, even though it seems difficult to define such a scale in a non-circular way. In contrast, ordering restrictions on what we have argued to be modifiers at the N^0 and N^1 levels seem inviolable—like the ordering restrictions between specifiers and modifiers, suggesting that the structural analysis proposed for these modifiers is on the right track. Something seems definitely degraded about (96) as compared with (95):

(96) ??I need the coffee plastic jar, not the tea plastic jar.

These considerations, as well as the stress and semantic considerations discussed throughout this paper have therefore led us to the conclusion that modifier-noun constructions can be either N^1 or N^0, no matter whether the modifier is adjectival or nominal. We now summarize the various classes of cases we have examined in this paper giving examples to fill out the

structural possibilities. Non-lexicalized [A N] combinations—instances of normal prenominal adjectival modification—are mostly phrasal. At least some of the productive cases of apparent N^0-level [A N] combinations may really involve [N N] modification at some level, as we argued in reference to (50), and as suggested by the discussion in Levi 1978. Lexicalized [A N] forms—those forms that must be listed since they have more or less idiosyncratic or idiomatic meanings—may be either N^1 or N^0:

(97)

	Lexicalized	Non-Lexicalized
N^0	**yellow** pages	**dental** operation
N^1	red **herring**	red **paint**

With [N N] modification we clearly get all possible combinations of Productive/Lexicalized and bar level. There is no sense that [N N] N^0 are more lexicalized than [N N] phrasal combinations:

(98)

	Lexicalized	Non-Lexicalized
N^0	**eye**ball	**Capsicum** leaf
N^1	rice **pudding**	dingo **stew**

Finally, we have noted compound-like examples where the modifier is a possessive. These may be either N^1 or N^0. In both cases the expression tends to be lexicalized, though there are some productive subsystems, such as those left-stressed examples headed by *milk*:[15]

(99) N^1 Swainson's **thrush**
N^0: **dog's** body

4.3 Some Other Compound-like Constructions in English and Elsewhere

We have discussed apparently left-headed nominals in English. On the basis of their stress behavior all such examples appear to be N^1. Many of these constructions have a questionable status in the minds of many speakers. Thus the prescribed plural *attorneys general*, which is what is expected given a left-headed structure, has already lost to the alternative *attorney generals*, suggesting that this form has been largely reanalyzed. Other examples, possibly because of their productivity, seem more solidly left headed: *Mounts Whitney and McKinley*. One can link these cases of post-nominal modification with the slightly marked, though still very productive class of post-nominal modifications with adjectives: *a man proud of his son*, *a river broad and long*, *a dragon fearsome to behold*.

Other compounds and compound-like words which do not fit so neatly into the discussion of most of this paper include exocentric bahuvrihi compounds. Bahuvrihi compounds have been analyzed (cf. Kiparsky 1982) as

[15]Note that some left stressed examples function as proper names and disallow preceding modifiers or specifiers in most cases, e.g., *Kayne's generalization*. There is nothing particularly surprising in this, and it seems reasonable to analyze such expressions as themselves being proper names, which typically lack articles in English.

modifiers to an empty headed noun (*redneck x*). Such an approach seems to solve the embarrassing problem that these constructions have no semantic head (a *redneck* is not in any sense a neck); however, we must assume in addition that these cases are all N^0, presumably a consequence of their having undergone at least the second stage of lexicalization outlined in Section 3.1

Examples like those of (35) are presumably nominalizations of phrasal fragments reminiscent of the Romance examples in (100):

(100) a. essuie-glace (wipe glass) 'windshield wiper'
b. tocadiscos (play records) 'record player'
c. guardaboschi (guard woods) 'forester'

Such examples are syntactic in appearance but have been transformed into nouns, suggesting a structure along the lines of that argued for in Di Sciullo and Williams 1987:

(101) $[_{N^0}\ [_{V^?}\ [_{V}$ toca] $[_{N^?}$ discos]]]

4.4 More on FCA-Only Theories: Ladd's Deaccenting Argument

We have taken the traditional generative view that syntax determines stress in the "normal" case—abstracting away from FCA effects. In adopting this view we are rejecting, among other approaches, Ladd's 1984 proposal that compound stress is a result of deaccenting the head. His idea is that the difference in representation between *steel warehouse* ('warehouse made of steel') and *steel warehouse* ('warehouse for storing steel') is purely metrical, the latter having the normal iambic phrasal structure *w s* and the latter the trochaic, deaccented structure *s w*. This deaccenting in compounds is explicitly claimed to be part of the more general phenomenon of phrasal deaccenting in examples like (102) (Ladd 1984, p. 255), where *books* is deaccented for pragmatic reasons:

(102) Has John read *Slaughterhouse Five*?
No, John doesn't **read** books.

Ladd's basic claim is that in right-stressed modifier-head nominal constructions, the attribute (lefthand member) does not serve to *subcategorize*[16] the head. In the deaccented (CSR) cases the attribute does subcategorize the head, hence the head contributes only part of what is necessary to identify the new category. An obvious contrast is *green* **house** versus **green** *house*. In the former the attribute *green* does not produce a new type of entity, but merely serves to further specify the general category *house*; the accented head therefore provides the category. A **green** *house*, on the other hand, is a new category of entity, so *house* provides only part of the necessary information, which results in its being deaccented relative to *green*, which

[16]Ladd is not using the technical linguistic sense of *subcategorize*, but intends the sense of "making a subcategory out of" the head.

provides crucial additional information. The approach is argued to work not only in cases where semantic categorization is involved, but also in cases where this is much less clear, such as in the domain of place names. So, the contrast between *First* **Avenue** and **First** *Street* is argued to derive from the fact that *Street*, within the category of names for roads, is the more common or expected term, and also gives the least semantic information about what kind of road is involved (terms like *Avenue*, *Boulevard*, *Alley*, carry with them connotations which *Street* does not have). Thus *Street* is deaccented.

While there are surely areas in which notions of subcategorization-conditioned deaccenting plays a role, the idea is problematic as a general explanation for the stress patterns we have seen. For example, one pragmatic test proposed by Ladd to distinguish "flavors" from real "(sub)categories" in food terms falls far short of correlating with stress. Ladd's idea is that *cheese* **sandwich** is a sandwich flavored with cheese and *cheese* merely serves to further specify *sandwich*, not form a separate category: thus the stress is phrasal. In contrast, for **banana** *bread*, one clearly needs the information provided by the lefthand member to determine the category of food involved since banana bread is not really a kind of bread flavored with banana, but a new category of food similar to bread in some ways. Now, Ladd notes that while one can felicitously ask *do you want a sandwich?* and subsequently offer a cheese sandwich, one cannot felicitously ask *do you want some bread?* and offer banana bread. It is claimed that one can only felicitously offer XY if XY is merely a Y flavored with X, and not a different category of food. This test is then supposed to correlate with stress, righthand stress if the offer is felicitous, lefthand if it is not, for the reasons given above. However, this test makes the wrong predictions in whole categories of examples. For instance, most words headed with bread are left-stressed even when their referent is a canonical instance: one can felicitously offer a person bread and then give them any one of *wheat bread*, *rye bread*, *white bread* and a number of others. On the other hand, while *steak and kidney* **pie** follows the normal pattern of righthand stress for words headed by *pie*, it would come as quite a surprise to most Americans to get some after assenting to an offer of unmodified pie.[17].

Another problem with Ladd's reduction of compound stress to general mechanisms of phrasal deaccenting is that phrasal deaccenting can easily produce examples of right-branching structures where all the righthand material is deaccented:

(103) I don't [**like** [buckwheat [griddle cakes]]]

Ladd's theory therefore predicts that left dominance in right-branching compounds is possible, if the subcategories fall out right. This does not

[17]This point can be made even more strikingly with *shepherd's* **pie**, a concoction wherein ground lamb is topped with mashed potatoes and then baked.

seem to be true, even where one might expect such stressing given the behavior of binary compounds with identical heads and lefthand members:

(104) a. **buckwheat** cakes
b. **griddle** cakes
c. buckwheat **griddle** cakes
d.*__buckwheat__ griddle cakes

Of course, (104d) is possible in FCA contexts, but it is not the default case for such compounds, contrary to what we might expect given Ladd's theory.[18]

4.5 Are N^0 Compounds Formed "in the Lexicon"?

In assuming, as we have been, that N^0 compounds are instances of modification constructions on a par with N^1 modification constructions, we have been at least implicitly assuming that such compounds are syntactic objects. Doing so, of course, runs counter to most theories of morphology, which assume that compounds are dealt with in the morphological component and that if compounds seem to share some properties with syntactic constructions this should nonetheless not be taken as evidence that they are syntactic (Selkirk 1982, Hoeksema 1984, Mohanan 1986, Di Sciullo and Williams 1987, among many others).

We would like to consider one recent attempt to justify this separation of compound formation from syntax, namely that of Di Sciullo and Williams 1987 (henceforth D&W), and consider whether the evidence they adduce for this separation is reasonable. D&W suggest an approach to morphology whereby words are constructed in a component of the grammar, the morphological component, which is in many ways similar to syntax, but is both (i) different in detail from syntax and (ii) separate from syntax. We discuss each of these points in turn.[19]

Differences of detail. A crucial difference of detail is that compounds are right headed in English, whereas phrases are left headed, at least at a very macroscopic level of analysis.

There are actually two issues relevant to comparing the headedness of phrases and compounds. One, D&W claim, is that one can always pick out the head in a phrase because it is "the item with one less bar level than the phrase" (p. 23) and which is generally of the same category as the phrase:

[18]Ladd's own examples involving the head noun *warehouse* might seem to be a contradiction to what we have just said since *warehouse* is, historically at least, a compound. However, it seems well within reason to assume that the analysis of *warehouse* as a compound is *only* of historical interest and that the word today lacks an internal word boundary. Certainly productively formed constructions with unequivocally complex heads do not seem to stress as Ladd ought to predict: a department store for electronics would be a *electronics* **department** *store*, except of course in clear FCA-contexts. Yet surely an electronics department store must be considered a subcategory of department store, at least if a steel warehouse is considered to be a subcategory of warehouse.

[19]See also Baker 1988 for a review of problems with D&W.

(105) $X^n \rightarrow \ldots YP \ldots X^{n-1} \ldots ZP \ldots$

They contrast this situation with the situation in a canonical compound where the two daughters of the N^0 are themselves N^0. They suggest that compounds follow a different, non-syntactic, morphological component rule for determining headedness, namely a modified version of the Righthand Head rule of Williams 1981, which states roughly that the head of a word is its rightmost morpheme. In compounds, it seems, the syntactic notion of head would not work.

Still, it is important to bear in mind that the ease with which one can apply the syntactic notion of head in phrases, and the seeming difficulty of using that notion in compounds derives from theoretical prejudice. To see this point, consider the structure of noun phrases in Welsh such as:

(106) meibion athrawon (sons teachers) 'teachers' sons'

This noun phrase consists of two plural nouns with no other markings, and yet there is no question that it is left-headed, as are all noun phrases in Welsh. In order to apply the syntactic notion of head, one would clearly want to argue that the righthand noun *athrawon* 'teachers' is really the sole member of a DP—hence could not count as the head under (105)—although there would be no way to ascertain the presence of the DP from the existence of *athrawon* alone. Returning to English compounds, there is no absolute reason why one cannot assert that all prenominal modifiers within N^0 are phrasal; we have already discussed examples which show that at least in some cases the prenominal modifier in the N^0 level must be an X^1. So, one might assume that *all* prenominal modifiers are X^1 just as we tend to assume that the noun *athrawon* in (106) is really a DP, despite the lack of direct evidence in the particular construction in (106). If we were to take that tack, we could clearly modify the syntactic notion of headedness to allow that the head of a construction must match the category of the parent and have a bar level not greater than that of the parent; something like this condition must be right more generally anyway if we are to define headedness correctly on adjunction structures under normal notions of adjunction. Then, the head of *dog house* would be clear: the prenominal modifier *dog* would be an N^1, whereas *house* would be an N^0, thus satisfying the conditions for headedness under the revised definition.

Still, we don't have to adopt this strategy. The other issue, raised by D&W's claim that English phrases are left headed is the question: left-headed at what level? Clearly, *modifiers*, whether phrasal or not, precede their heads in the typical case in English, and since we have assumed that N^0 compounds are merely a type of modification, we would expect that compounds are right-headed. Thus in order to determine the head, the syntax would merely need to take note of the kind of relationship—i.e., modification—holding between the members of the compound, and fix the head on the right. D&W, in claiming that English phrases are left headed,

have missed the locally more relevant point that in modifier-head constructions in English the head is on the right.

Of course, in deriving compound headedness in this way, we fail to link the right-headedness of compounds with the right-headedness of words in English: for example, many people (e.g., see Williams 1981, and also Lieber 1980) have argued that the head of, say, *grammaticality* is the suffix *-ity* since that determines the category. Williams' Righthand Head Rule explicitly links compounds with affix right-headedness. We think, in contrast, that the two should not be linked, at least synchronically. The fact that it is typically suffixes and rarely prefixes that function as heads of words is almost surely related to the cross-linguistic prevalence of suffixation, which in turn may have psycholinguistic bases (see Cutler et al. 1985). Thus it seems plausible to view the right-headedness of English compounds and the right-headedness of English affixation as a coincidence, in the sense that they are not diverse expressions of a single fact about English grammar.[20]

Indeed, we would expect languages with left-headed modification structures by and large to also show left-headed compounds if they have compounding, no matter what the headedness of affixation is. This appears to be correct, as examples from French, discussed in Selkirk 1982, or from Spanish, Breton "loose compounding" (Stump 1989) or Zhuang (Ramsey 1987).[21] For further discussion of this and related points see Lieber 1988, 1992, Baker 1988, and especially Clements 1989:

(107) a. timbres-poste (stamps post) 'postage stamps'
b. los coches Chevrolet
c. Breton: kaoc'h kezeg (dung horse) 'horse dung'
(cf. paotr brav (boy fine) 'a fine boy')

[20]Of course, this is not to say that all other patterns would have been equally likely, or that the two facts about English are not connected historically. As Cutler et al. 1985 show (see, e.g., pp. 727ff), there is a positive cross-linguistic correlation between suffixing and *phrasal* right-headedness. In particular, right-headed languages—Object-Verb and Postpositional languages—show an overwhelming preference to be exclusively suffixing. Left-headed languages mostly have both prefixes and suffixes, and prefix-only languages are very rare and are apparently always phrasally left-headed. English descends from languages which were phrasally right-headed, as did Latin, from which half of English affixational morphology comes. English prehead modification is likely to be a living fossil of the former right-headed syntax. So right-headed compounding may be historically related to right-headed affixation in English, by a history which makes crucial excursions into the syntax.

[21]We should note that Welsh, which has left-headed modification structures, has a number of frozen right-headed compounds: *ysgolfeistr* (school-master) 'school master'; *rheilffordd orsaf* (rail road station) 'railway station'. Some examples, such as the second one, are clearly translation borrowings from English, and in any event this kind of compounding is not productive in Welsh. Breton "strict compounds" (Stump 1989) are also right-headed. Note that Welsh (and also Irish, Breton, French and Spanish) does retain a few prenominal adjectival modifiers and so it may well be the case that Celtic and Romance languages are only somewhat further along the transition to left-headedness than is English.

d. Zhuang: kai5-pou4 (chicken male) 'rooster'
(cf., sai1 mo5 rau2 (book new our) 'our new book')

Concerning the French examples, D&W claim (p. 83) that "Selkirk misanalysed fixed phrases such as *timbres-poste* as 'left-headed compounds,' thus concluding that French morphology is a mixture of left- and right-headed structures (since affixation in French is clearly right-headed). This is a clear example of a generalization compromised by the failure to properly separate syntax and morphology."

This attack can easily be turned around, however: by insisting that morphological constructions, including compounds, must obey the Righthand Head Rule, D&W make it impossible both to recognize as compounds many constructions which any first-order description would classify as such, and to state the very clear generalization that compounds in a language generally follow the ordering of modifier-head constructions in that language. In the other direction, relying on the Righthand Head Rule as a correlate of "morphological" constructions would prevent us from noting properties that left-headed examples in other languages may share with English compound nouns. Thus in cases like *carne de burro* (meat of donkey) 'donkey meat' in Spanish, it is essential that the complement phrase *de burro*, like the modifiers in English compounds, not be a full DP. With an article, the phrase *carne del burro* 'meat of the donkey' would not serve to refer to the generic substance donkey meat. *Carne de burro*, while having clear phrasal properties, equally clearly has properties commonly associated with compounds.

Separation from syntax. The other half of D&W's claim about the status of compounds is based upon the apparent *atomicity* of words from the point of view of syntax. In saying that compounds are atomic, D&W are buying into the familiar claims of lexical integrity (see, e.g., Mohanan 1986, pp. 24–25) which say that the syntax (or subsequent components) can have no access to the internals of words. To substantiate this claim, D&W present evidence that appears to show that various syntactic phenomena fail to make use of information internal to compounds. The strength of the evidence, however, fades rapidly on serious examination.

Consider, for example, the observation (D&W, p. 49) that rules of syntax cannot "discriminate among compounds whose nonhead members are plural or singular," so that *the parks commissioner* is treated by all relevant syntactic rules exactly the same way as *the park commissioner*. This seems simply to be an observation about headedness: in exactly the same way, relevant syntactic rules will not distinguish *the commissioner of meats* from *the commissioner of meat*. Since plurality is not being registered on the head of the noun phrase in either case, any syntactically relevant rule which cares about number (such as verb agreement) will register the grammatical number of the head and not be affected by the plurality of modifiers

or complements. In this respect there is absolutely nothing special about the compound case: one does not need to appeal to the existence of a separate morphological component to derive the relevant properties of *the parks commissioner*.

A second set of examples concerns cases like **it robber* (see D&W, p. 50) or **who killer*, which are argued to show that referential expressions, which are relevant to the syntax, may not occur within words. But, as we have already suggested (Section 4.1), there seems to be a constraint ruling out any kind of derivation from closed-class items; alternatively, one could claim that no projection of functional categories can serve as modifiers. In any case, we have argued that modifiers—whether at the N^0 or the N^1 level—are maximally X^1, thus ruling out a whole class of DP or DegP modifiers. These considerations subsume D&W's facts. Referential expressions are not in principle ruled out from occurring within words, but only those expressions which have the above mentioned forms. Proper names can certainly occur within compounds although D&W suggest that: (i) only famous names really allow this, since **Bill admirer* is supposedly unacceptable; and that (ii) even in such cases as *Nixon admirer*, *Nixon* doesn't really refer to Nixon. We briefly examine these claims.

With respect to the second claim, it has been argued extensively in Sproat and Ward 1987 and Ward et al. 1991, that names which are morphologically contained within compounds can indeed be used to refer to particular individuals, and this can be seen by the fact that such names can function as antecedents to pronouns:[22]

(108) a. There's a Thurber$_i$-story about his$_i$ maid...
b. I was reading this Peggy Noonan$_i$ book on her$_i$ years at the White House...
c. We went up to Constable$_i$ country; we stayed in the village he$_i$ was born in.

D&W's own evidence for their claim concerning the referentiality of *Nixon* in *Nixon admirer* is the assertion that in (109a), admiring Nixon is not an essential property of a *Nixon admirer*, since this example is not a contradiction, unlike (109b):

(109) a. John is a Nixon admirer in every sense except that he does not admire Nixon.
b. John admires Nixon in every sense except that he does not admire Nixon.

Ward et al. (1991, footnote 16) have argued that (109a) is only non-contradictory under the interpretation of *Nixon admirer* as a person with a reliable set of traits (such as being clean-shaven, wearing a three-piece suit and always carrying an attaché case) which are independent of the

[22]Contrary to the predictions of the so-called *Anaphoric Island constraint* (Postal 1969).

property of admiring Nixon: in other words, *Nixon admirer* would have to have accreted additional meanings via institutionalization, along the lines of *Maoist*. In that case a person might be said to have those properties and thus qualify as being a *Nixon admirer*, even if he or she does not actually admire Nixon; hence (109a) would not be a contradiction. Needless to say, *Nixon admirer* has not achieved this institutionalized status in the minds of most speakers, and the contrast in (109) is therefore very hard to get.

More important for D&W's argument, however, is the observation that the putative accretion of such institutionalized meanings is largely orthogonal to the issue of whether *Nixon* in *Nixon admirer* can be used to refer to Nixon. Hence, even making the counterfactual assumption that the contrast in (109) were a clear one, that would not serve to support D&W's claim concerning the referential properties of compound-internal elements.

Returning to point (i), the obvious question which arises when confronted with an example like *Bill admirer* is: who is Bill? One might suppose that *Bill admirer* ought to be able to mean something like 'one who admires people with the name *Bill*', but there is really no reason to believe that it should have this reading: in its typical phrasal usage, a proper name like *Bill* is invariably used to refer to a specific individual. Therefore one would expect *Bill* in *Bill admirer* also to be used to refer to some specific individual, and if we do not know who that individual is, the example seems odd. The oddity is reinforced by the fact that, as is well-known, "agentive" nominals tend to have a "habitual" reading, and are typically used to refer to characteristic properties. To ascribe to someone the characteristic property of admiring some unspecified person named *Bill* seems odd indeed. Naturally, famous names do not have this problem, since there is no question about the intended referents of the names. And *Bill admirer* is, it seems to us, fine in a context where the interlocutors can assign a definite referent to *Bill*.

Still, there is no question that first names generally seem somewhat degraded over last names in such contexts: compare *?Noam admirer* with *Chomsky admirer*. At least in part this seems to have to do with the fact *admirer* has an institutionalized connotation of 'admires in the professional sense', and this in turn requires that the admired individual be referred to using their normal professional designation, typically the last name.[23] But familiarity or conventional designation are clearly extragrammatical considerations, and there is no reason to suppose that there should be a statement in the *grammar* of English nominal compounds restricting the occurrence of proper names in that context.

[23] A similar constraint applies to cases of deferred reference (thanks to Mats Rooth for the second example):

i. You'll find Chomsky on the top shelf.

ii. ?#You'll find Noam on the top shelf.

In summary, we do not find any of the reasons adduced by D&W for separating compounding from phrasal constructions to be compelling. Therefore it seems possible to accept the assumption that English N^0 compounds are word-level objects that are formed in the syntax.

5 Two Approaches to Parsing and Stress Assignment

We turn now from the descriptive and theoretical issues which have occupied us for the bulk of this paper to the issue to which we would ultimately like to provide a solution, or at least a good approximation thereto, namely the correct parsing of and stress assignment to modified nominals as they occur in unrestricted text. We outline here some of the methods we have applied to solving this problem.

It is of practical concern that the correct stress pattern for a sequence $[N\ N]$ or $[A\ N]$ cannot be generally predicted from the preterminal sequence of categories alone. That is, if we know that we have a sequence of two nouns, we cannot be certain that we have an N^0 as opposed to an N^1 for a parent category, though if the preterminal sequence were all we could compute, guessing N^0 would more often lead to a correct stress assignment than guessing N^1. In general, some other type of analysis is necessary. In addition to the problem of deciding upon the appropriate stress assignment to a binary modified noun, there is the at least equally difficult problem of parsing more complex cases. As we shall see, the two problems, stress assignment and parsing, are related in the methodologies which one can apply to them. We shall first turn to the problem of stress assignment, and then consider parsing.

5.1 Stress Assignment in Binary Nominals

The phrase *safety board* has main stress on the left. As we have argued for other cases in the text, we might decide this because the phrase is an instance of a semantic pattern, informally stated as *N1 GROUP-WHOSE-CONCERN-IS-N1*. In Sproat and Liberman 1987 and Sproat 1990, we show that a rather crude implementation of such a method can improve considerably on the performance of an approach based purely on the syntactic categorization of the words in the text stream. However, as we noted in our discussion of the semantic patterns of argument-argument compounds, it is not at all clear on what basis a single, consistent, coherent set of such schemata can be defined.

Alternatively, we might depend on the fact that the words in the phrase *safety board* tend to yield a left-dominant stress pattern in other cases. So one might expect that the probability that the nominal will be stressed on the left given that the first word is *safety* and the second *board* is greater than the probability of righthand stress under those conditions, or in other words:

$$p(s_1|w_1{=}safety)\,p(s_1|w_2{=}board) > p(s_2|w_1{=}safety)\,p(s_2|w_2{=}board)$$

A survey of examples from a large corpus of naturally-occurring pre-modified nominals shows that this is indeed the case—*safety* as a noun modifier was always stress-dominant, while *board* when pre-modified by a noun was never stress-dominant. A representative sample:

(110) a. safety belt, safety binding, safety council, safety devices, safety equipment, safety feature, safety glasses, safety harness, safety margin, safety measure, safety net, safety nut, safety precautions, safety problem, safety procedure, safety razor, safety record, safety regulations, safety requirement, safety rule, safety school, safety shoes, safety standards

b. Federal Reserve board, advisory board, bulletin board, chalk board, chess board, circuit board, cutting board, dart board, draft board, drawing board, governing board, ironing board, memory board, particle board, planning board, promotion board, school board, score board, skate board, supervisory board, surf board, wall board, zoning board

Although most examples are not as clear-cut as this one, we believe that the method can work fairly well if properly trained. Its main drawback is that many words do not occur often enough in the needed constructions to generate useful statistics—it seems appropriate, in such cases, to depend on the observed behavior of "similar" words. The crux of the matter is then the similarity metric to be used.

5.2 Parsing Premodified Nominals

Again, we have explored two methods. One approach would use phrasal schemata of the kind used in assigning stress to decide upon a possible analysis for a modifier-head string; one might further extend such a method along the lines of Finin 1980 by adding scores for the various schemata, and pick the parse whose cumulative score is highest. This approach has the previously noted difficulties of creating a reasonable set of schemata, and, if scores are used, also the problem of juggling the interaction of possibly ad hoc scores, endemic to diagnostic systems of this type. Nevertheless, if the schemata are chosen so as to reflect the patterns found in a certain genre of text, the technique can work reasonably well.

A system of this kind currently under development in the context of Bell Labs ongoing text-to-speech effort is reported in Sproat 1990,[24] and is an extension of earlier work reported in Sproat and Liberman 1987. The program depends on semantic patterns of the kind described in this paper as well as large lists of common binary nominals to attempt to compute

[24]The scoring method alluded to above for picking among various possible semantic interpretations is not implemented in the system reported in Sproat 1990, though various ad hoc heuristics are used to pick one of several possible analyses.

a reasonable structure of multiply modified nominal phrases. As a simple example consider a semantic grammar which contains information that *table*, *chair* and some other items are *HOUSEWARE*s; that *kitchen*, *bathroom* and *living room* are *ROOM*s (and incidentally also N^0, this latter information being crucial for the binary cases); and that there is a pattern, which says that a *ROOM* word and a *HOUSEWARE* word may be combined into an N^1, which we will call a *ROOM&THING*. this rule can be thought of as an instance of the pattern *PLACE-WHERE-N2-IS-FOUND N2* discussed in (70). In addition to semantic grammar rules and the rather large lexicon of binary cases, there are more traditional syntactic rules which handle syntactic aspects of noun phrase structure, such as the treatment of determiners. The system in its current form can assign appropriate structure and stress to noun phrases such as the following:

(111) $[_{DP}\ [_{POSS}$ John's$]\ [_{NP}$ large $[_{\langle N^1,\ ROOM\&THING\rangle}\ [_{\langle N^0,\ ROOM\rangle}$ living room$]\ [_{\langle N^0,\ HOUSEWARE\rangle}$ table$]]]]$

In this example, *living room* is correctly analyzed as a modifier of N^1. Another module of the program responsible for assigning stress can then take this assigned structure and determine that the main stress should be placed on the head noun in this case.

The second approach to parsing again relies on the statistical behavior of individual words, identifying the affinity of two adjacent words with the extent to which they occur together more often than one would predict based on how often they occur individually. One reasonable measure for such affinity might be MUTUAL INFORMATION, where the mutual information $I(a,b)$ between events A and B with probabilities $p(a)$ and $p(b)$ is defined as

(112) $I(a,b) = \log_2 \frac{p(a,b)}{p(a)p(b)}$

Mutual information measures have been used recently by Sproat and Shih (1990a)[25] in a domain rather similar to the current one, namely the problem of locating word boundaries in Chinese text. Chinese orthography, of course, traditionally does not indicate the location of word boundaries, but it turns out that considerations of the strengths of association between adjacent characters in a Chinese text, as measured by mutual information can achieve about 95% retrieval and precision for two-character words (which constitute the bulk of multicharacter words in Chinese text).

However, there is a problem with defining the association measure strictly in terms of mutual information as defined above. Instances of the sequence ABC in fact give us no information about the relative affinity of B for A as opposed to C, so if a significant fraction of (say) AB instances occur in ABC sequences, while BC is quite a bit more common outside

[25]See also Magerman and Marcus 1990 for more extensive use of mutual information in parsing.

A	B	C	AB			BC			ABC
F	F	F	F	I	IX	F	I	IX	F
interagency	*task*	*force(s)*							
35	756	6909	4	10.83	00	287	10.07	**10.06**	4
environmental	*impact*	*statement(s)*							
801	1110	4626	25	8.40	**7.50**	16	5.23	3.25	12
private	*arms*	*dealer(s)*							
3296	3822	1059	21	4.32	4.02	57	7.40	**7.30**	4
electoral	*college*	*system(s)*							
405	1949	4691	93	10.47	**9.29**	62	9.23	2.04	59

Table 1

these sequences, then a high value for I(A,B) will be misleading. It is better to compute a mutual-information-like affinity measure that excludes the useless ABC instances from the counts. If we use $-T$ to refer to the complement of the cases in which T is found, our affinity estimate for words A and B within the ternary unit T becomes:

$$(113)\quad IX_T(A,B) = \log_2 \frac{p(A,B \mid -T)}{p(A \mid -T)\,p(B \mid -T)}$$

Now, compare the right-branching ternary nominals *interagency task force* and *private arms dealer* with the left-branching ternary nominals *environmental impact statement* and *electoral college system*. In a 12-million-word sample of the *Associated Press* newswire, things work out as in Table 1.

In this example, the IX measure[26] gives the correct analyses, while the I measure fails for the case of *interagency task force*, which has the noted property in that the sequence *interagency task* never occurs other than in the larger frame. In general, the IX measure gives a stronger indication of affinity in the correct direction, even where the I measure is also correct.

Methods of this general type have a great deal of promise as aids to parsing in the all-too-common cases where structural indications are weak or lacking. However, it would be desirable to consider, in estimating the affinity between two words, the distribution of "similar" words.[27] Also, the generalization of the simple ternary case to more elaborate structures can be done in a variety of ways, and it is not trivial to find methods that are both tractable and correct. We believe that the analyses given in this paper will help to guide such explorations in sensible and productive directions.

[26]Note that we are using *00* to refer to the case where a pair does not occur other than in the triple.

[27]Indeed, one of the problems encountered in using this statistical measure as an additional option in the system reported in Sproat 1990 is that there is often not enough data about particular pairs of words in a given triple to estimate IX, even if one uses a moderately large corpus of about 10 million words.

References

Abney, Steven. 1987. *The English Noun Phrase in its Sentential Aspect.* Doctoral dissertation, Massachusetts Institute of Technology.

Baker, Mark. 1988. Morphological and Syntactic Objects: A review of Di Sciullo and Williams' *On the Definition of Word. Morphology Yearbook* 1:259–283.

Bauer, Lauri. 1983. *English Word-Formation.* New York: Cambridge University Press.

Blåberg, O. 1988. *A Study of Swedish Compounds.* Doctoral dissertation, Umeå University.

Bolinger, Dwight. 1967. Adjectives in English: Attribution and Predication. *Lingua* 18:1–34.

Bolinger, Dwight. 1972. Accent is Predictable (if you're a mind-reader). *Language* 48:633–644.

Chomsky, Noam. 1970. Remarks on Nominalization. In *Readings in English Transformational Grammar*, ed. Roderick A. Jacobs and Peter S. Rosenbaum, 184–221. Waltham, Massachusetts: Ginn and Company.

Chomsky, Noam, and Morris Halle. 1968. *The Sound Pattern of English.* New York: Harper and Row.

Clements, J. 1989. Lexical Category Hierarchy and Syntactic Headedness in Compounds. In *Proceedings of the 6th Annual Eastern States Conference on Linguistics*, 46–57.

Cutler, Anne, John Hawkins, and Gary Gilligan. 1985. The Suffixing Preference: A Processing Explanation. *Linguistics* 23:723–758.

Di Sciullo, Anne-Marie, and Edwin Williams. 1987. *On the Definition of Word.* Cambridge, Mass.: MIT Press.

Downing, Pamela A. 1977. On the Creation and Use of English Compound Nouns. *Language* 53:810–842.

Dowty, David. 1979. *Word Meaning and Montague Grammar.* Dordrecht: Reidel.

Fabb, Nigel. 1984. *Syntactic Affixation.* Doctoral dissertation, Massachusetts Institute of Technology.

Finin, Timothy W. 1980. *The Semantic Interpretation of Compound Nominals.* Doctoral dissertation, University of Illinois at Urbana-Champaign.

Fudge, Erik. 1984. *English Word-Stress.* London: George Allen and Unwin.

Hayes, Bruce. 1980. *A Metrical Theory of Stress Rules.* Doctoral dissertation, Massachusetts Institute of Technology (distributed by IULC).

Hirschberg, Julia. 1990. Using Discourse Context to Guide Pitch Accent Decisions in Synthetic Speech: The given/new distinction and deaccentability. In *Proceedings of AAAI-90*, 952–957.

Hoeksema, Jacob. 1985. *Categorial Morphology.* New York: Garland Publishers.

Jackendoff, Ray. 1977. *X-bar Syntax: a Study of Phrase Structure.* Cambridge, Mass.: MIT Press.

Kiparsky, Paul. 1982. Lexical Morphology and Phonology. In *Linguistics in the Morning Calm*, ed. I.-S. Yang, 3–91. Seoul: Hanshin.

Ladd, Robert. 1984. English Compound Stress. In *Intonation, Accent and Rhythm*, ed. Dafydd Gibbon and Helmut Richter, 253–266. Berlin: Mouton de Gruyter.

Lees, Robert B. 1960. The Grammar of English Nominalizations. *International Journal of American Linguistics, Publication 12.*

Levi, Judith N. 1978. *The Syntax and Semantics of Complex Nominals.* New York: Academic Press.

Levin, Beth, and Malka Rappaport. 1992. *-Er* Nominals: Implications for a Theory of Argument-structure. In *Syntax and Semantics 26: Syntax and the Lexicon*, ed. Eric Wehrli and Tim Stowell. New York: Academic Press.

Liberman, Mark, and Alan Prince. 1977. On Stress and Linguistic Rhythm. *Linguistic Inquiry* 8:249–336.

Lieber, Rochelle. 1981. *The Organization of the Lexicon.* Doctoral dissertation, Massachusetts Institute of Technology.

Lieber, Rochelle. 1983. Argument Linking and Compounds in English. *Linguistic Inquiry* 14:251–286.

Lieber, Rochelle. 1988. Phrasal Compounds in English and the Morphology-Syntax Interface. In *Papers from the 24th Regional Meeting of the Chicago Linguistic Society*, 398–405. University of Chicago.

Lieber, Rochelle. 1992. *Deconstructing Morphology: Word Formation in Syntactic Theory.* Chicago: University of Chicago Press.

Magerman, David, and Mitchell Marcus. 1990. Parsing a Natural Language Using Mutual Information Statistics. In *Proceedings of AAAI-90*, 984–989.

Maidment, J. A. 1989. On the Accentuation of Compounds in English. In *Speech, Hearing and Language: Work in Progress*, 181–188. University College, London.

Marantz, Alec. 1989. Projection vs. Percolation in the Syntax of Synthetic Compounds. Manuscript, University of North Carolina.

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

Mohanan, K.P̃. 1986. *The Theory of Lexical Phonology.* Dordrecht: Reidel.

Postal, Paul. 1969. Anaphoric Islands. In *Papers from the 5th Regional Meeting of the Chicago Linguistic Society*, 205–239. University of Chicago.

Ramsey, S. Robert. 1987. *The Languages of China.* Princeton: Princeton University Press.

Roeper, Thomas, and Muffy Siegel. 1978. A Lexical Transformation for Verbal Compounds. *Linguistic Inquiry* 9:199–260.

Safir, Ken. 1987. The Syntactic Projection of Lexical Thematic Structure. *Natural Language and Linguistic Theory* 5:561–602.

Selkirk, Elizabeth. 1984. *Phonology and Syntax.* Cambridge, Mass.: MIT Press.

Selkirk, Elizabeth O. 1982. *The Syntax of Words.* Cambridge, Mass.: MIT Press.

Sproat, Richard. 1985. *On Deriving the Lexicon.* Doctoral dissertation, Massachusetts Institute of Technology.

Sproat, Richard. 1990. Stress Assignment in Complex Nominals for English Text-to-Speech. In *Proceedings of ESCA Workshop on Speech Synthesis*, 129–132.

Sproat, Richard, and Mark Liberman. 1987. Toward Treating English Nominals Correctly. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, 140–146. Association for Computational Linguistics.

Sproat, Richard, and Chi-lin Shih. 1990a. A Statistical Method for Finding Word Boundaries in Chinese Text. *Computer Processing of Chinese and Oriental Languages* 4:336–351.

Sproat, Richard, and Chi-lin Shih. 1990b. The Cross-linguistic Distribution of Adjective Ordering Restrictions. In *Interdisciplinary Approaches to Language*, ed. Carol Georgopoulos and Roberta Ishihara, 563–593. Dordrecht: Kluwer Academic Publishers.

Sproat, Richard, and Gregory Ward. 1987. Pragmatic Considerations in Anaphoric Island Phenomena. In *Papers from the 23rd Regional Meeting of the Chicago Linguistic Society*, 321–335. University of Chicago.

Stump, Greg. 1989. A Note on Breton Pluralization and the Elsewhere Condition. *Natural Language and Linguistic Theory* 7:261–273.

Thomason, Richard. 1985. Some Issues Concerning the Interpretation of Derived and Gerundive 'Nominals'. *Linguistics and Philosophy* 8:73–80.

Ward, Gregory, Richard Sproat, and 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* 12:245–274.