ROBERT B. LEES

PROBLEMS IN THE GRAMMATICAL ANALYSIS OF ENGLISH NOMINAL COMPOUNDS

Ten years ago, in a study of English nominal expressions, I attempted to contribute to the long and still continuing research on compounding.[1] At that time I conceived of the problem in the following terms: in English, German, Turkish, and other languages, the stock of nouns in the lexicon may be extended indefinitely by the creation of composite nouns, each characteristically of two members, themselves already *bona fide* words or compounds, the whole, a so-called NOMINAL COMPOUND, pronounced with the unifying stress-pattern typical of single-word nouns.

It is easy to see that such compound expressions are of indefinitely great internal complexity, for the members of a nominal compound may themselves be composite:

Pûre Fôod and Drúg Làw
ânti-Viêt Nâm wár dêmonstràtor
vítamin defíciency disêase
súgar càne plantâtion ôwner

Thus, since ideally there can be no longest nominal compound in the language, *i.e.*, since the set of all well-formed nominal compounds of English must contain (countably) infinitely many members, one must suppose that a speaker's knowledge of WHICH composite nouns count as well-formed expressions of his language must be formulable only in terms of some grammatical RULES. My problem was to specify what rules suffice for the generation of English nominal compounds.

Since the grammatical structures which these rules would assign to the generated compounds must contribute essentially to a speaker-hearer's understanding or interpretation of each compound, I attempted to formulate the structure of each type of compound so that the syntactic relation between its parts would reflect that which occurs in some grammatically underlying expression, ultimately an independent sentence. For example, the internal organization of the compound *dráwbrìdge* can be

[1] Lees, R. B., *The Grammar of English Nominalizations*, Indiana Univ. Research Center in Anthropology, Folklore, and Linguistics, Publication No. 12, 1960; esp. Chapt. IV. [Abbrev. herein *GEN*].

assigned by construing it as a transform of the nominal expression *bridge for someone to draw*, itself a reduction of the relative-clause construction:

bridge which is for someone to draw

and thus ultimately a derivative of a nominal which contains an embedded sentence

Someone draws the bridge

Thus, the formal relation between the two parts *draw* and *bridge* was said to be just the syntactic relation between those two words in this underlying sentence, that is, the relation of a verb to its direct object. This compound is, of course, not an isolated instance, but it is drawn from a productive set: *blówpìpe, fláshlìght, púsh-bùtton sétscrèw, stópwàtch, tóuchstòne, wásh-drèss, etc.* Each is understood in terms of this Verb/Direct Object relation.[2]

The study revealed compounds of many internal structures and included the syntactic relations among the functional sentence-parts Subject, Predicate Nominal, Possessive Object, Verb, Direct Object, and Oblique (prepositional) Object. Different derivational rules might enumerate compounds of contrasting form but ones which reflect the same underlying syntactic relations. Thus, parallel to the infinitival *dráwbridge* type there would also be the gerundial type of *chéwing gùm, drínking wàter, réading matêrial, smélling sàlts, wéaring appàrel, etc.*, which also reflect the Verb/Direct Object relation.[3]

Since the rules proposed for the analysis of compounds can be validated only if they comprise a functional part of a generative grammar of the language as a whole, and since such a grammar must provide primarily a description of whole sentences, the rules which generate nominal compounds not only synthesize them from their parts, but they must also serve to position each compound described just in those places inside of sentences where they may properly be used. For example, the compound *grówing pàins* can be used in a sentence only where *plural* nouns may occur; similarly, *lócksmìth* is an *animate* noun, *stéambòat* inanimate, but *chátterbòx* is again animate. Thus, the rules always apply to both a so-called MATRIX-SENTENCE within which a given compound functions as a noun and also to an embedded sentence, the source of the compound's parts.[4]

A critic has recently pointed out that on this view the assignment of syntactic relations to certain compounds seems ambiguous.[5] In some, one member of the compound appears to be derived grammatically from an element of the embedded sen-

[2] These compounds were all described as the endocentric, infinitival V-O type of SETSCREW, in: Lees (1960: 152).

[3] Described as *for*-adverbial V-O type of EATING APPLE (from: *apple for eating*) in: Lees (1960: 153).

[4] Called "constituent sentence" in: Lees (1960: 153ff).

[5] C. Rohrer, Review of *GEN, Indogermanische Forschungen* 71 161 (1966).

tence, while in another type of compound the corresponding member appears to be derived from an element of the matrix-sentence.

Or conversely, for example, *éating àpple* is described as a transform of the nominal expression

...apple which is for eating...

and the compound *clótting àgent* as a transform of

...agent which causes clotting (of something)...

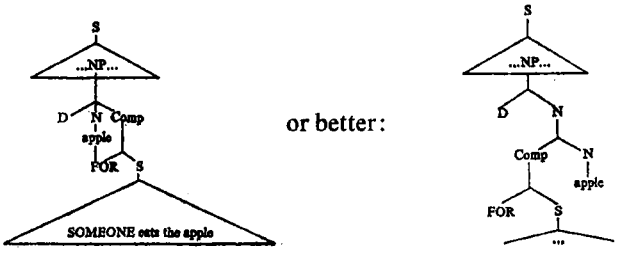
In both cases the head-noun in the compound appears to be a transform of a gerundial element in the embedded relative-clause sentence; yet *éating àpple* is characterized as reflecting the Verb/Direct Object relation, *clótting àgent* as exemplifying the relation of Direct Object/Subject![6]

There is indeed some ambiguity in our use of such concepts as Subject, Object, *etc.*, for they refer now to the so-called grammatical subject or object, now to the so-called 'logical' subject or object. In sentences containing several echelons of successive embedding only a complex diagram can make clear the reference of these grammatical terms.

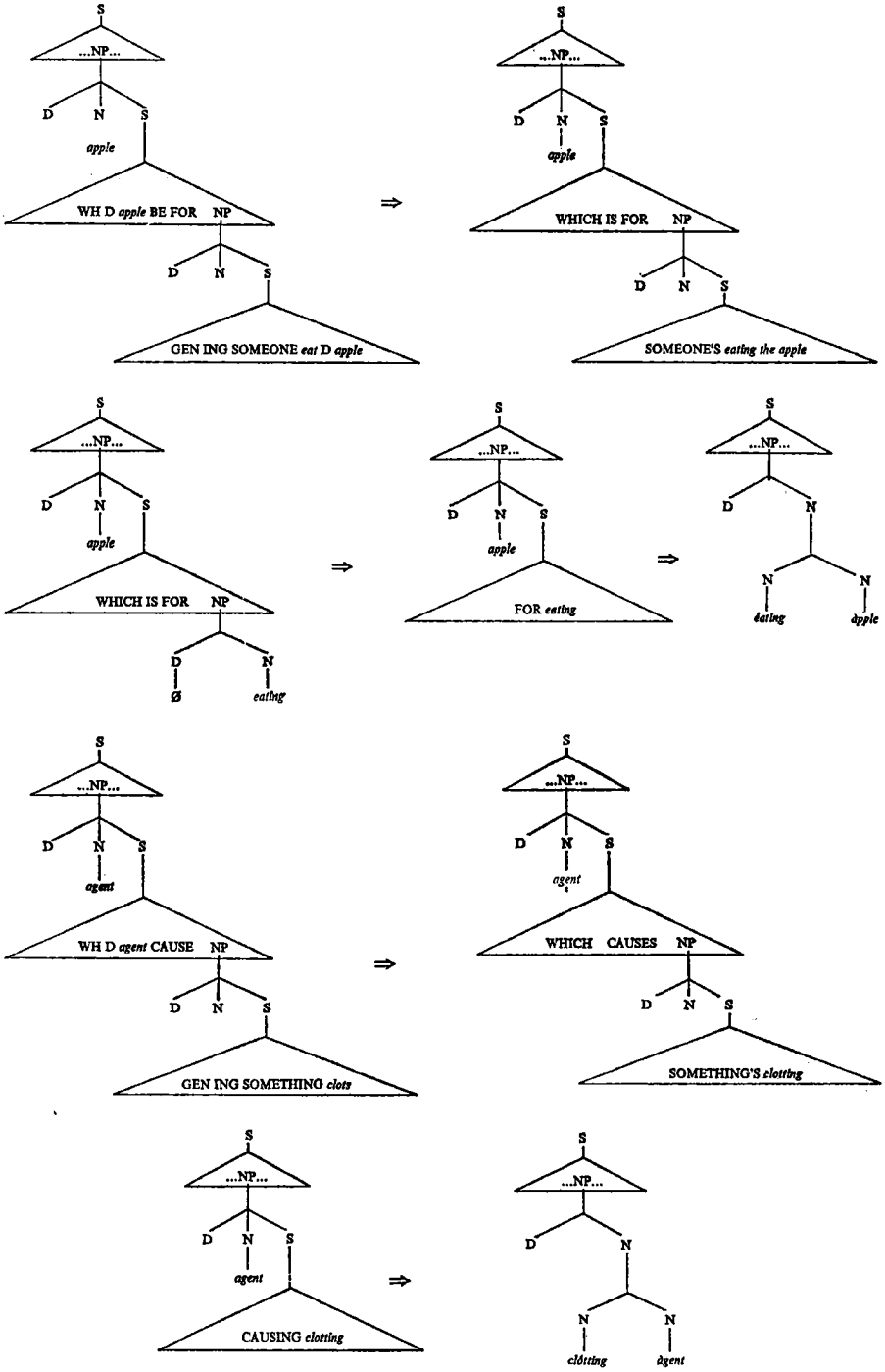
In other words, we must distinguish clearly between the superficial structure of expressions and their deeper syntactic organization, perhaps even at several levels of depth. The syntactic relations which the members of compounds were said to reflect are, of course, those definable only over the innermost embedded expressions underlying the compounds. If we represent this deeper syntactic structure abstractly in the form of a branching-diagram of constituency, then the contrast between the *éating àpple* and the *clótting àgent* cases becomes quite clear:[7]

[6] Described in *GEN*, p. 147, as of a small subclass of the O-S type of CAR THIEF in which the first member happens itself to be a gerund: *láughing gàs, sléeping pìll, snéezing pòwder*, etc.

[7] In these schematic derivations I make no strong claim about the detail of the transformational relation between the deepest structure and the compound itself. Thus, the underlying form might well be:

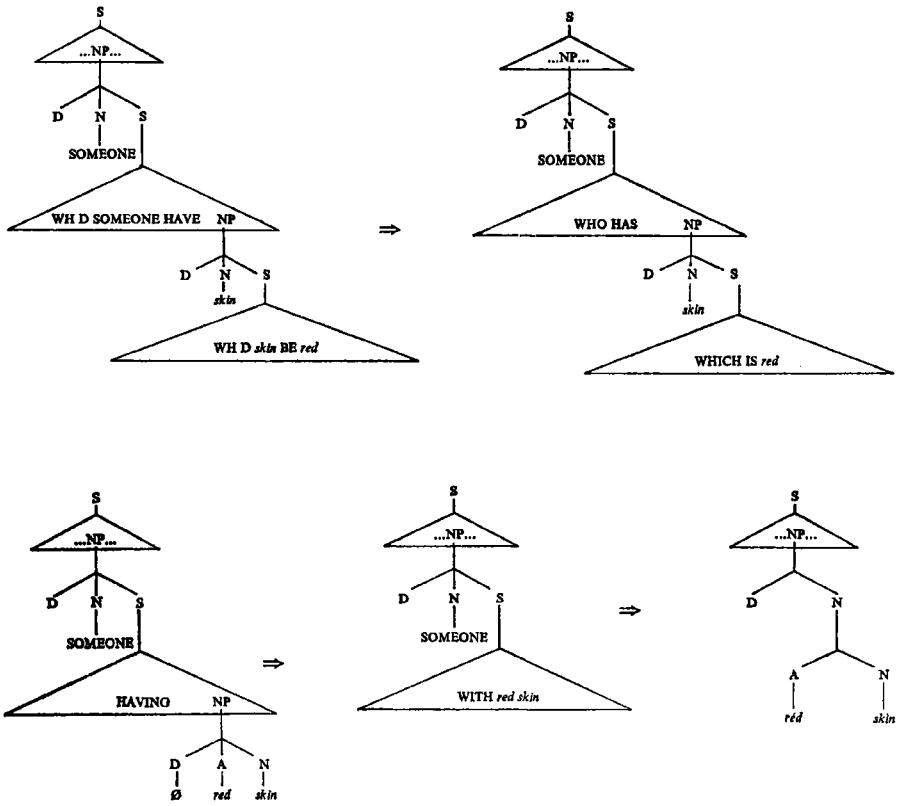


complicating the base-component of the grammar slightly but simplifying the transformational correspondingly.



Now, simply classifying compounds according to the deeper syntactic relations reflected by their constituents is clearly not the same thing as providing an ANALYSIS of their syntactic structures. We can illustrate this well with another example of two kinds of exocentric compounds typified by *rédskìn* and *prónghòrn*.[8] The constituents of the first reflect the relation between the Predicate Adjective and the Subject of a copula sentence, while those of the second have the relation of the two nominals in a sentence of the form: NP_1 *resembles* NP_2. Yet the rest of the analysis is much the same for the two.[9]

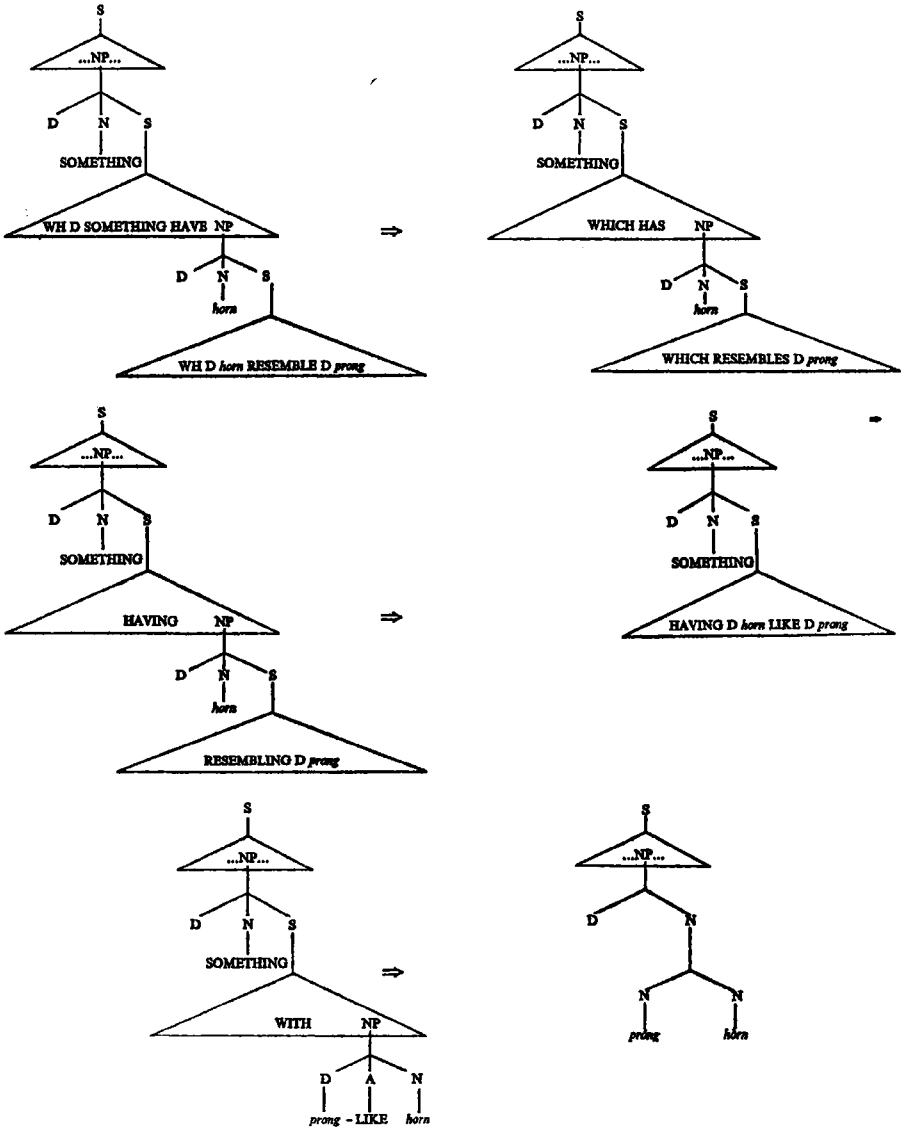
rédskìn 'someone who has skin which is red'



[8] Also discussed by Rohrer in his review, Sec. IV.

[9] Described resp. as the exocentric, adjectival S-Pred. Nom. type, *GEN*, p. 130, and the exocentric, O-like-S, copulative, S-Prep. O type of EGGHEAD, p. 159.

prónghòrn 'something which has a horn which resembles a prong'



The efficacy, or explanatory power, of these analyses depends, of course, upon the generality and plausibility of the rules required to project from these abstract constituent-trees in each case to the pronounced compound expression itself, the surface-structure.[10]

[10] E.g., the major compound-generating rule of *GEN* was a generalized form of the rule T-86: $N_1 + P + N_2 \rightarrow \bar{N}_2 + \bar{N}_1$, pp. 133, 173-5.

The critic mentioned above has also pointed out another peculiarity of such grammatical analyses and has thereby raised a much more serious question. It certainly seems reasonable to construe the syntactic relation reflected by the constituents of certain compounds to be that of Subject and Direct Object, to choose one type in question. Now, on that view, the verb of the underlying sentence must in the analysis somewhere be deleted by a general rule of ellipsis. For example, if we analyze the compound *cár thìef* as reflecting the syntactic relations of a sentence *The thief steals cars*, as seems reasonable, then from the tree underlying that sentence, when it is embedded as a compound inside of a matrix-sentence, the verb *steal* must be deleted. But if the compound *bédbùg* is analyzed similarly as reflecting the syntactic relations of Object and Subject in a sentence *The bug infests beds*, then the verb *infest* must be deleted.

However, a problem now arises in our conception of how such a grammatical analysis might underlie a hearer's interpretation of an arbitrary compound of this type. Either we must assume (1) that he understands every such compound to be as many ways ambiguous as there are different verbs which could have served non-deviantly in the relevant deep-structure sentence, or we must assume (2) that the grammar somewhere associates explicitly with each Object/Subject pair just the right verbs which may be deleted in the formation of the compounds allowed.

Alternative (1) would mean that a person might, within the rules of well-formedness of English, construe the compound *bédbùg* as a version not only of *bug which infests beds*, but also of such expressions as *bug which steals beds*, *bug which eats beds*, *etc.*, though not, say, of such deviant expressions as **bug which encourages beds* or **bug which purchases beds*. The non-deviant cases correspond to 'legal' constructions of English which may happen not to be in current use. Alternative (2), on the other hand, would mean that a very restricted set of verbs, possibly only one, would underlie each such compound, that they would thus always be known to the listener, and that the differences in verbal meaning among individual compounds in this class are correctly construed either (a) as part of a speaker's NON-linguistic knowledge or (b) as contained within the semantic description of the head-noun itself. Thus, illustrating case (a), we might suppose that if there is a difference in the verbs underlying *áshtràу* and *bírdcàge*, it is due only to our linguistically independent knowledge of the difference between what we do (in our culture) with ashes in a tray and what we do with a bird by means of a cage; illustrating case (b), we might suppose that the difference between *góldsmìth* and *bráin sùrgeon* is exactly reconstructible from the meanings of *smith* and *surgeon*.

In my opinion it is not easy to decide between these two alternative conceptions. But a more accurate and detailed analysis of compound expressions may yet render the conflict obsolete.

Now, for a number of good reasons several contemporary grammarians have been led to view the syntactic relations of Subject and Object as relatively superficial

features of sentences.[11] In other words, the only difference between, *e.g.*, an active sentence and its corresponding passive version is the choice of whether the agent noun will be the 'topic of conversation' or the patient noun. Thus, the deepest representation of the structure of the sentence must contain choices for AGENT and PATIENT, not subject and object. The same superficial choice of topic may also distinguish the members of such sets of related sentences as:

(a) *I have an alligator in my room*
My room has an alligator in it
There is an alligator in my room

(b) *We filled the tank with tangerine juice*
The tank was full of tangerine juice
Tangerine juice filled the tank

(c) *They used a laser to carve the steak*
They carved the steak with a laser
It took a laser to carve the steak

From these and other considerations we are led to construct a deeper syntactic form for sentences directly in terms of noun-phrases which fulfill certain universal semantic functions, such as (human) agent, (human) patient, (non-human) instrument, location, time, purpose, or mere general, unspecified participant.[12] To these we must add a classification of verbs according to which of these noun-phrases they may obligatorily or optionally select as complements, as well as a set of transformational rules to map deep-structures onto surface-structures. These rules include those which choose from among the noun-phrases under specified conditions which will become Subject, which Object, *etc.*

Reapplied to the interpretation of nominal compounds, such an analysis permits a much finer distinction and a closer connection between the meanings of a compound and its deep syntactic structures. It may also afford a decision on the problem of indiscriminate verb ellipsis mentioned before.

For example, re-analyzing a large class of Subject/Object compounds including *wíndmìll* and *háy fèver*, we may isolate a productive set with the underlying structure:

Verb General Complement Noun Instrument Noun

that is, with the verb-phrase structure V-O-I. (In sentences with this selection of noun-phrases, normally the I-noun becomes Subject, the O-noun the Object.) But these compounds all fall into two subsets in such a way that the V of each member of one subset may be viewed as a variant of the verbs *impel*, *energize*, *activate*, *power*, *propel*, *etc.*, while the V of each of the others is a variant of *cause*, *engender*, *produce*,

[11] E.g., C.J.Fillmore, J.Ross and G.Lakoff, J.Lyons, and M.A.K.Halliday.
[12] This version of the proposal to deepen our picture of underlying syntactic structure is due to Fillmore (1968).

yield, etc. Thus, in the first set we find *wíndmìll, stéambòat, hýdrogen bòmb*, and in the second set *háy fèver, néttle ràsh, báttle fatìgue, etc.*

In other words, it may be possible to associate one, or a small number of generalized verbs with certain classes of compounds by fixed grammatical rule, so that the compounds in question need not be described by the grammar in such a way as to imply that they are indefinitely ambiguous. By 'generalized verb' I mean just the minimal set of semantic features which characterize all variants in the sets: *impel, propel, energize, activate, power, drive, actuate, etc.*, or: *cause, engender, produce, yield, ...*

The case cited is not isolated. Another example arises within the class of compounds re-analyzed as exhibiting the syntactic relations of sentences whose verb-phrases contain:

Verb General Complement Agent Locative,

that is, have the form V-O-A-L. Again there are two subsets whose verbs seem to be semantically characterizable. In one class, containing *bírd càge, pígpèn, dóg hòuse, sált cèllar* (!), *etc.*, the verbs are all variants of whatever underlies *keep, nurture, store, confine, house, etc.* In the other, including compounds such as *téa ròom, grócery stòre, lúnch còunter, banána pòrt, etc.*, the verbs are variants of whatever underlies *sell, deal in, service, handle, etc.*

It is by no means clear yet that all cases of compound types which contain as members two of the nouns of a verbal sentence can plausibly be analyzed in terms of only a few generalized verbs. Such compound types must include at least the following:

(1) $\text{V-O-A} \rightarrow \text{N}_2\ \text{V-s}\ \text{N}_1 \rightarrow \acute{\text{N}}_1 + \tilde{\text{N}}_2$

$\text{N}_1\ \text{N}_2$

áirplàne pîlot	gúnsmìth
bríck màson	hórse dòctor
cár thìef	pástry chèf

[These, as suggested above, require no special assumptions about the underlying verb since the latter is reconstructable from the meaning of the head-noun.]

(2) $\text{V-O-I} \rightarrow \text{N}_2\ \text{V-s}\ \text{N}_1 \rightarrow \acute{\text{N}}_2 + \text{N}_1$

$\text{N}_1\ \text{N}_2$

(a) V = *energize, drive, power, actuate, propel, impel, ...*

áir rìfle	óil stòve
álcohol làmp	stéambòat
héat èngine	súction pùmp
hýdrogen bòmb	wáter whèel
mótor càr	wíndmìll

(b) V = *cause, yield, engender, emit, produce, ...*

báttle fatìgue	fínger prìnt	vírus disèase
blóod stàin	háy fèver	
cándle lìght	ínk blòt	
cóld sòre	néttle ràsh	
díaper ràsh	sáw dùst	
féver blìster	sóap sùds	

[Described above]

(3) V-O-A-I → AV-s N_1 with N_2 → $\acute{N}_1 + \grave{N}_2$

N_1	N_2

(a) V *repel, prevent, reject, forestall, suppress, remove, ...*

búg sprày	héadache pìll
cóugh sỳrup	líghtning ròd
fíre èngine	mosquíto nèt
flý pàper	móthbàll
gás màsk	ráin càpe

(b) V = *preserve, ensure, protect, retain, foster, secure, ...*

chástity bèlt	lífebòat
chícken wìre	sáfety lòck

(c) V = *provide, vend, supply, afford, produce, ...*

cóke machìne	sóap òpera	wáter tòwer
eléctron gùn	tóne àrm	
pícture tùbe	wáter pìstol	

(d) V = *determine, measure, establish, ...*

defléction gàuge	distórtion mèter	hóur glàss

(e) V = *exhibit, portray, show, ...*

fáshion shòw	flówchàrt	wánt-àd

(f) [V as in (a)]

búg sprày	íon tràp
héat shìeld	mosquíto nèt
ínsect repèllant	nóise fìlter

[Perhaps reconstructible from head-noun, as in (1)]

(4) V-O-L → N_1 V-s in N_2 → $\acute{N}_2 + \grave{N}_1$

N_1	N_2

V = *live, work, (in)fest, (in)habit, ...*

(a) animate endocentric

bédbùg	cáve màn

bóll wèevil fíeld mòuse
bánk tèller hóspital òrderly stóre clèrk
garáge mechànic párlor màid

(b) exocentric
gróund hòg práirie dòg wáter mòccasin
gúttersnìpe séa-hòrse

(c) inanimate
bódy flùids hóuse dùst spáce chàrge
gróund wàter kídney stòne

(5) V-O-A-L → A V-s N_1 in N_2 → $\acute{N}_1 + \grave{N}_2$
| |
N_1 N_2

(a) V = *keep, nurture, put, raise, ...*
áshtrày dóg hòuse insáne asỳlum
bírd càge dóvecòte mádhòuse
bríef càse dúck pònd rábbit hùtch
búll pèn flówer bèd sált cèllar (!)
cár-bàrn gréenhòuse shéepfòld
cówshèd ícebòx sick bày

(b) V = *sell, deal in, service, ...*
banána pòrt frúit màrket téaròom
bárgain còunter grócery stòre
béauty shòp hámburger jòint
bíble bèlt lúnch còunter
bórscht círcuit mílk bàr
cóffee hòuse stóck màrket

(6) V-O-A-L → A V-s N_1 in N_2 → $\acute{N}_2 + \grave{N}_1$
| |
N_1 N_2

V = *use, put, ...*
áircràft móuthwàsh táble wìne
cóffee crèam nósedròps
cúrry pòwder pócketbòok
éyewàsh sándwich sprèad
fíeld artìllery séaplàne
fóotwèar schóol gràmmar
gárden pàrty shóe pòlish
háirbrùsh spaghétti sàuce
hándcùff stáge mòney

(7) V-O-A-Ab → A V-s N_1 from N_2 → $\acute{N}_1 + \tilde{N}_2$

N_1 N_2

(a) V = *get, obtain, derive, ...*

cóal mìne	límestòne	rúbber trèe
fúr sèal	músk dèer	sílkwòrm
grável pìt	óil wèll	stóne quàrry
hóney bèe	ópium pòppy	súgarbèet

(b) V = *make, prepare, concoct, ...*

ápple sàuce	grápe wìne	wóod àlcohol
béet sùgar	óatmèal	
blóod sàusage	péanut bùtter	
chéese sprèad	potáto chìps	
cóal tàr	róotbèer	
físh càke	róse wàter	

There are, of course, still other large classes of compounds which may not reflect the syntactic relations of noun complements of a verb but rather those of the constituents of certain copula sentences, or those of the genitive constructions, *etc.* The following illustrative examples are a few of the types not yet fully investigated:

Object/Property	cóllar sìze, vápor prèssure
Whole/Part	árrowhèad, cártwhèel
	óyster shèll, ónionskìn
	cátgùt, whále bòne
Contents	múdhòle, pícture bòok
	sándbàg, áir pòcket
	gróund wàter, kídney stòne
Resemblance	búlldòg, háirsprìng
Form.	bríck chèese, lúmp sùgar
	fírebàll, ráindròp
Material	tínfòil, páper mòney
	bútter còokie, shórtning brèad

More sophisticated analysis in the study of compounding provides some evidence for the view that the deepest syntactic structure of expressions is itself a more or less direct picture of their semantic descriptions!

In conclusion let me try to anticipate and meet a possible objection to this view. After struggling very hard to escape the obscurities of mentalistic philosophizing in the study of language, late 19th-century grammarians adopted a more and more rigidly empiricist doctrine on scientific rigor and validity in linguistics. This doctrine prescribed that an acceptable description of a language could contain nothing but direct generalizations from the spoken or written expressions themselves. Abstract,

theoretical, or intervening variables were greeted with suspicion. Such a methodology left very little which could be said about the relation between sound and meaning. Its results contributed correspondingly little to our understanding of how a language user formulates what he wishes to say or interprets what others say to him. We have not yet entirely overcome the excesses of that empiricist or behavioristic period of development in the social sciences.

Thus, there may be some who would view the suggestions I have made, or have reviewed here, as an unfortunate return to pre-scientific classifying of expressions according to their supposed meanings, or to an undesirable semantically-based grammar. Others might feel that to re-iterate eternal verities, while not reprehensible as such, is hardly a contribution to our linguistic knowledge.

But contemporary grammatical analysis is not a mere re-discovery of the obvious or of what our ancestors have bequeathed to us. Scientific description is a continuing effort to render ever wider and deeper bodies of observations perspicuous and natural to our understanding. Replacing an older theory by a slightly improved modern one enables us to bring within its scope a larger variety of otherwise unconnected and *ad hoc* phenomena.

Accordingly, the value of these suggestions on compounding lies mainly in the way they serve to relate older and largely correct insights into universal semantic categories and deep syntactic structures to a wide spectrum of independently analyzed expressions within the framework of a unified conception of linguistic organization.

REFERENCES

Lees, Robert,
1960 *The Grammar of English Nominalizations* (= *Indiana University Research Center in Anthropology, Folklore, and Linguistics, Publication* No. 12).
Rohrer, Christian,
1966 Review of Lees (1960), *Indogermanische Forschungen* 71:161.
Fillmore, Charles J.,
1968 "The Case for Case", in *Universals in Linguistic Theory* (New York, Holt, Rinehart and Winston, forthcoming).