The Generic Vocabulary and The Opacity of Compounds[1]
Edwin Williams
March 12 2021

Di Sciullo and Williams (1987) (D&W) proposed and defended the principle of the "atomicity" of words, including compound terms. As formulated there, atomicity, a variant of the "Lexicalist Hypothesis", dictates that

(1) "the inside of X^os is inaccessible to syntactic (or sentence-level semantic) rules"
     (Atomicity Condition, D&W p. 49)

where by X^os was meant words, including compounds[2]. They cited, among other things, Tense and pronouns as elements that could not appear inside of compound terms (p. 50) on the grounds that compounds were atomic in phrasal syntax.[3] In this note I will look at new evidence for the D&W proposal for compound terms, leaving aside for the moment the more general question of words.

But I will try out a different notion of atomicity here, one which I will call "opacity". Rather than blocking rule-access to compounds, as in (1), the opacity condition will instead insist that

(2) A compound cannot contain a variable free in it
     (Compound Opacity Condition (COC))

For many cases this gives the same result as the D&W definition of atomicity; for example, if a movement rule leaves a trace, the trace is a free variable in every expression containing it, up to the phrase that contains the moved constituent. An equivalent variant of (2) would be to say that compounds must be "assignment independent", in the sense that they must contain no elements which vary under the Tarskian "assignment of values to variables" by which semantic interpretation is composed. So, to use the example from D&W, there can be no "it robber" (where "it" refers to some bank) because "it robber" has a free variable in it, and would get different semantic values in different contexts because of that variable.

I think that opacity so understood can in fact explain a larger class of cases in which compound terms fail than was envisaged in D&W. These include pronouns and Tense as in D&W; definite and indefinite NPs; contextually interpreted items like *different*, *local*, and the like; speech-event indexicals like *now* and *today*, reflexives (apart from a special class); and operators with complements like *more-than*. In addition, there are a surprising number of cases of valid compounds that look like they should be excluded by both the D&W atomicity condition and COC, but on more refined analysis turn out not to be excluded, and COC permits the more refined analysis. For example, the rule of scope assignment (QR) is barred by D&W atomicity from applying not only *into* compounds, but also *within* compounds;

---

2   The use of the term "syntactic" here led to confusion. Instead of "syntactic rule" D&W should have said "rule of phrasal syntax", as the formulation in (1) suggested to some that D&W did not think that words had "syntax", despite the fact that their book was devoted to laying out a syntax of words and its relation to phrasal syntax.
3   See Anderson (1992) for a nearly identical formulation in his "Principle of Lexical Integrity"

but, if the scope of a quantifier is entirely contained within a compound, then the variable it binds complies with COC, and such cases will be discussed in this note.

An important theme of this note is that it seems unlikely that COC can be inferred from or built into the rules or principles for constructing compounds in the first place; rather, it appears to have the character of a filter applied to already-generated structures. The discussion of this point occurs at the end of section 4 and in the concluding section 7.

Wherever possible in the discussion I will contrast the compound construction with the possessive construction. In some ways they live side by side—we can say almost synonymously either "the hospital entrance" or "the hospital's entrance"; but on the point of the kind of opacity at issue here they are entirely different (but see section 7).

In the following sections I will look into the larger class of cases that fall under COC. At the end I will raise the question of why the opacity condition exists—what purpose does it serve as part of the language definition.

1. Pronouns, Tense, and Proper Names

1.1 Pronouns

Strictly speaking, the D&W atomicity condition cited above is not quite enough to rule out "it-robber", despite the claim made there. The question is, what rule of grammar does the atomicity condition block in the case of a pronoun inside a compound? There is Binding Theory, but making a pronoun inaccessible to Binding Theory means only that Condition B will not apply to it, not that it will be ungrammatical. And the assumption alive then, as now, is that there is no rule of grammar to assign antecedents, or referents, to pronouns; grammar excludes some possibilities via Binding Theory (plus Weak Crossover), and whatever remains is allowed by grammar. So "it robber" survives.

This is the conceptual reason to prefer COC in terms of "free variables" or "variation under assignments" to the D&W atomicity condition—it assimilates elements not assigned interpretation by grammar to those that are. "it robber" is ruled out because it contains a free variable, *it*. But in addition there is an empirical advantage: sometimes a pronoun is in fact not a free variable, and in those cases the pronoun can occur in a compound:

(1)   a. he-man
       b. me-generation
              (Deschaine and Wiltschko (2002))
       c. she-wolf
       d. me too movement

In these cases the pronoun does not get, or need, a referent, as it plays a different role, and does not vary in meaning under different assignments to variables; so these occurrences of pronouns do in fact satisfy COC.

1.2 Proper Names

D&W tried to assimilate proper names to pronouns, I think wrongly. They pointed out that a term like "Nixon admirer" could be applied to people who did not literally admire Nixon (p. 50), suggesting that the proper name did not have its usual semantic value. Even if the observation is true, the interpretation is probably wrong. In fact proper names occur routinely in compounds, in sharp distinction to pronouns:

(2)   a. The Johnny Carson Show
      b. The George Miller Prize
      c. The UK Freedom Foundation
      d. The AstraZenica vaccine
      e. The Rockefeller dynasty

And they appear to have their normal semantic value. Suppose that when Jay Leno took over the Johnny Carson Show they had left the name of the show unchanged. "Johnny Carson" would still refer to Johnny Carson, not to Jay Leno, it's just that Johnny Carson's relation to the show would not be "person who hosts", but now something different.   This brings to mind Kripke's (1980) analysis of the role of "Godel" in "Godel's Incompleteness Theorem", and the conclusion is the same: not only are proper names not like pronouns, they are not like definite descriptions either, which also cannot occur in compounds (at least not as top-level terms of a compound and never in an assignment-dependent sense; see section 6). It looks like proper names are, as Kripke called them, "rigid designators", which simply means that they are assignment-invariant, and so have no free variable, and so are not like pronouns at all, and so occur in compounds freely.

Some proper names do vary under assignment—"Miss America", for example (Dowty, Wall and Peters (1981 p. 176). Even these show up in compounds, but with a sharp restriction on their meaning, in fact the COC restriction:

(3)   The Miss America Show

In such a case as (3), "Miss America" is not used in its shifting sense; although the referent of "Miss America" varies from year to year, the referent of "The Miss America Show" does not vary from year to year (at least, it does not vary as a function of who is Miss America; if the 1966 and 1967 Miss America shows were different shows, they would have still been different even if the same person won both years).

Significantly, the varying sense of "Miss America" cannot be used in a compound:

(4)   a. I visited Miss America's hometown
      b. I visited the hometown of Miss America
      c. I visited the Miss America hometown
      d. I read the Jane Austen book

In (a) and (b) "Miss America" can be used in its shifting sense, and might mean I visited the hometown of the current Miss America. But in (c) this is impossible; if (c) means anything, it means I visited a town generically associated with the title "Miss America" in some way, not the hometown of any

3

particular person; compare to (d), which means I read a book individuated in part by its relation to the individual *Jane Austen.*   (10a) also shows the clear difference between compounding and the possessive construction, in line with the expectations that COC gives rise to.

1.3 Noun Incorporation

I think that acknowledging that proper names can generally be included as terms in compounds leads to a sharpening of the notion of "genericity" that we want to understand.

Although I will not treat Noun Incorporation in detail in this note, I would hope that COC could be extended to it. There is reason to be optimistic about that—Sapir (1911) goes to great pains to distinguish incorporated nouns from cliticized pronouns (251 ff); since COC makes exactly this distinction—barring pronouns and allowing nouns—his discussion is good evidence for submitting Noun Incorporation structures to COC, either by subsuming them under compounding, or in some other way. And since COC tolerates proper names, it is of further significance that there are Noun Incorporating languages that incorporate proper names (Johns 2007).

In addition, Noun Incorporation, like compounding, can sometimes incorporate a modifier along with the noun:

(5)   a. [new car] [showroom]
         b. Biilersualiorsimavoq   [Inuktitut]
             biili- suaq- lior -sima -voq
             car   big   make -PST -3SG.S
             'He has made a big car.'   (Sadock 1980 p. 303, cited in Barrie and Mathieu 2016 p. 31)
         c. ngii-gchi-g2wiiwzensiw   [Ojibwe]
             n- gii- gichi- gwiiwzens -w
             I – PST-big-boy -be.VAI
             'I was a big boy'   (Barrie and Mathieu 2016, p. 18)

Furthermore, Noun Incorporation can never incorporate a determiner or quantifier, a further point of similarity with compounds, as will be discussed in section 3.

With these points in mind, we may bring to bear the findings about Noun Incorporation on the COC and the question of "genericity". Sapir (1911) determines that Noun Incorporation occurs in two distinct uses, "characteristic" and "accidental" (p. 264), which we may provisionally understand to corresond to "generic" and "non-generic", the distinction we are trying to understand.   I suggest that we identify Sapir's two uses with the two modes of Noun Incorporation in D&W (p. 63 ff)[4] One mode involves existential closure of the incorporated argument. The other involves "qualification" (D&W's term) or "classification" (Mithun's (1983) term) of the incorporated argument, but without satisfaction of it; that argument may then be satisfied by a full NP argument exterior to the incorporation structure.

---

4   See also Mithun's (1983) 4 classes of Noun Incorporation, the first three of which correspond to D&W's "satisfaction" type.

We may then notate Sapir's two modes of Noun Incorporation in the following way (for hypothetical "rabbit-eat"):

(6) a. $\lambda x \to Ey$ (eat (x,y) & rabbit (y))   ("characteristic", or saturating incorporation)
     b. $\lambda(x,y) \to$ eat (x,y) & rabbit (y)   ("accidental", or classifying incorporation)

In neither case is there a free variable that COC would object to; in (a) the y variable is bound within the incorporation structure; in (b) the y variable is a heritable property of the head (see the discussion of "head accessibility" at the end of the next section), and so immune to COC (as is the x variable in (a)). Saturating incorporation would then give rise to Sapir's "characteristic" class, and classifying incorporation to his "accidental class".

So, what kind of thing is Noun Incorporation, from a semantic point of view? Identifying the instances of compounding and of Noun Incorporation of the sort discussed in this note as terms that refer to "natural kinds" is much too narrow—natural kind terms should not include proper names, at the very least.   As Chierchia (1998) says, "Lexical nouns identify kinds. Complex nouns may or may not". "Picture" is a kind, "big picture" might be a kind, "picture of Sue" surely is not. Chierchia says as well that what counts as a kind is "set by the shared knowledge of a community of speakers". Compounds certainly fail this measure in general, and Noun Incorporation as well, if we consider, for example, the following examples from Mithun (1983):

(7) a. ... bene-dulg-nay mangaralaljmayn.   [Gunwinggualso]
            they.two-tree-saw cashew.nut
            '... They saw a cashew tree.'
     b. ... bene-red-nay redgereyeni.
            they.two-camp-saw camp.new
           '... They saw a camp which was freshly made.' ('They saw a new camp.')
                    Mithun 1983 p. 867

It is hard to imagine that "tree seeing" or "camp seeing" are items in the "shared knowledge of a community of speakers". Mithun makes the point that the incorporated nouns in languages like Gunwinggualso (her type IV, the classifying kind) are general nouns like *tree*, and not specific nouns like "cashew tree", but that is not the same thing as saying that the N+V combinations are kinds—*tree* is a kind and *seeing* is a kind but "tree seeing" and "camp seeing" are probably not. Mithun coins the term "nameworthy", which I think might correspond to Chierchia's notion of kind, but applies it only to the first (or possibly first three) of her four categories of Noun Incorporation, so it does not characterize NI structures in general, nor of course English compound structures. And lastly, as mentioned earlier, there are languages that incorporate proper names, and I would assume that such structures would be neither kinds, nor natural kinds, nor characteristic, nor nameworthy, nor generic.

I would suggest then that COC captures exactly the right notion of genericity for English style compounding and for at least the outer bounds of Noun Incorporation—no sensitivity to context.

The other terms may be useful for distinguishing different styles of compounding or Noun Incorporation. Or for characterizing the possible meanings of morphemes—I assume that the meaning of a morpheme cannot reference a proper name, or reference the referent of a proper name (without itself being a proper name).   From this perspective "fox hunting" is "nameworthy" enough that there

5

could be a morpheme with the same meaning (chasing down foxes on horseback with dogs but without guns); and in fact, in some communities in Virginia and Tennessee, that is exactly what the morpheme *hunt* means. But there could be nothing equivalent if *fox* were replaced by a proper name. Of course there is a lot to argue about here concerning what is meaning and what is "real world knowledge"—like, does the meaning of the verb "to bork" reference Robert Bork?

1.4 Tense

The important insight about Tense that we will exploit here is that Tense is a relation between times, not a reference to a time (see e.g. Reichenbach (1947) and all subsequent literature). It is a relation between the event-time of the clause in which the Tense occurs and some other time—either the event-time of the clause one up from it (for complement clauses) or the time of the speech event (for main clauses and relative clauses). As a relation it has two variables; one is bound to the event-time of the immediate clause to which it belongs, the other is bound later, as just described; it is this second variable that gives rise to COC violations in compounds.

(8)   a. John said$_{T:past}$ [he was$_{T:past}$ leaving]
        b. John will$_{T:present}$ say [he was$_{T:past}$ leaving]
        c. John saw$_{T:past}$ the woman [who was$_{T:past}$ arrested

In (a) and (b) the embedded Tense gets its second value from the matrix time reference—it is understood to be coincident with it or prior to it, as that is the past Tense relation in some languages, including English. The matrix Tense gets its second value from the utterance event. In (c) the embedded past Tense gets its second value independent of the matrix time reference, and so can have any relation to it, as both time references are fixed by the utterance event, but independently of one another.

Given that Tense always gives rise to a variable that is free in the clause in which it occurs, we should expect that Tense would behave like pronouns and essentially not occur in compounds, and this is almost correct:

(9) a. *John's that he was leaving soon remark turned out to be false.
    b. John's remark that he was leaving soon turned out to be false.

But there are two important carve-outs. The first is the tense equivalent of "he-man" cases—where the item in question does not contribute its normal semantic value to the phrase containing it. One instance where this arises is with quotation; we see this in the following:

(10)   John said "I am leaving soon"

Here the normal sequence of tense restriction—which forbids present in the complement of a past tense verb—is lifted, and in fact the embedded present is not fixed by the utterance event of the whole sentence either. It is context independent. And in fact, to a noticeable degree, quoted clauses can occur in compounds much better than normal complement clauses:

(11) ?John's "I am leaving soon" remark turned out to be false.

6

While not perfect it is much better than (3a).

This thinking extends to a variety of fixed expressions that can be used in compounds, like "a never-say-die smile". D&W thought that complex phrasal inclusions in compounds was limited to these fixed quotations and hyphenated expressions. See section 7 for a demonstration that the phrasal parts of compounds are fully productive up to the limits of COC.

The second, and more important, carve-out has to do with tensed compound verbs, like "pan-fried" "hand-made", etc., The Tense in such cases does get its second value from the context of the compound as a whole, as a function of either the event-time of a higher clause or of the utterance event:

(12)   a. I pan-fried the fish
        b. I thought I had pan-fried the fish
        c. [pan [fry   ed$_{T:past}$]]

So, why isn't this a violation of COC, since Tense occurs inside of a compound in (c)? The crucial point is that Tense in (c) is the head of the compound, and like the head of any word or phrase, it is accessible from outside the compound. There are various mechanisms to bring this about. D&W suggested that by virtue of X-bar syntax an expression had all the properties of its head, including its ultimate head, and so "T:Past" became a top-level property of [pan-fried], and so was accessible to sentence and utterance level rules, via inheritance through the *head-of* relation. A probably more popular approach is to have Tense generated outside of the morphological verb altogether, let it be interpreted there, and then later "affix hop" it onto (or here, into) the morphological verb, as in the 1957 *Syntactic Structures* model, and Distributed Morphology. I will assume here that some mechanism can accomplish the "accessibility of the head" that is needed here without choosing among the different possibilities.

Whatever the mechanism turns out to be, it maybe should be extended to pronouns as well:

(13) John plays two different roles in the play, both animals.
     He knew that audiences were disgusted by dog-him, but found cat-him irresistible.
                        (scenario suggested by R. Kayne p.c.)

To the extent that "dog-him" and "cat-him" are acceptable, and they are way better than "it robber", they show that the pronoun can have its full semantic value from inside a compound that it is head of—in (7) "dog-him" unmistakably refers to *John*.

2. Quantifiers

        Quantifiers are the original case of variables (the traces of quantifier expressions) getting assigned values. Since the trace of the quantifier cannot be free in a compound, a quantifier must take its scope within the compound to comply with COC—that is the cold clear prediction. I can't possibly do justice to the variety of natural language quantifiers in this note, but I will look at some sharp comparisons that involve the different scope properties of different quantifiers to judge how well those differences align with COC.

2.1 Universal quantifiers

It has been widely observed that *all* and *every* differ in one scope-related property—*all* can be interpreted *in situ*, in which case it gives a "group" interpretation, but *every* must be assigned nontrivial scope:

(1) a. Every friend of John's thinks that they are his best friend
     b. All John's friends think that they are his best friend
     c. All John's friends think that they are his best friends

(b) has a meaning equivalent to (a), but the meaning of (c) cannot be paraphrased with *every*. This is the "group interpretation" of *all*, not available for *every*. Since it doesn't take scope, it doesn't allow pronouns to get a distributed interpretation like *every* does.

If that is so, then we might expect to find *all* in compounds, but not *every*, and this appears to be correct:

(2) a. a flour for every purpose
     b. a flour for all purposes
     c. an all purpose flour
     d. *an every purpose flour

There is one loophole—suppose *every* could take scope over the entire compound in this way:

(3) [every purpose$_i$ [t$_i$ flour]]$_N$

Then COC would be satisfied so long as we did not subject the constituent [t$_i$ flour] to it. But it is likely that N is not a valid scope domain for *every* anyway, not being "propositional".

2.2 Numerals in adjunct vs. argument position

We start with the facts:

(4) a. a 6 drawer chest
     b. *a 3 car owner
     c. a 3 car family

An obvious difference between (b) and (c) is that "3 car" is an argument of *owner*, but not of *family;* and "6 drawer" is not an argument of *chest*. So, it looks like a numeral in an argument position in a compound is ungrammatical. Given the previous section, we might speculate that numerals in argument position must get scope, but that in adjunct position, they need not.   This makes conceptual sense, since, at least in clauses, a quantified expression in adjunct position is already adjoined to a possible scope for itself, a sentence-like domain, whereas in argument position, or at least in direct object position, it is not adjoined to a possible scope.

What counts as an adjunct here? The following leads to a strange conclusion:

(5)   a.   a 1000 round magazine
        b. *a 1000 round holder
        c.   a 1000 round shell holder

(a) and (b) are clear enough—"1000 round" is not an argument of *magazine*, which takes no arguments, but it is an argument of *holder*. The surprise is (c), which is intended as synonymous with (b), but nevertheless escapes COC.   Apparently, *shell* is the argument of *holder* here, and "1000 round" counts as an adjunct[5], so (c) reduces to (a).

A further puzzle is "1000 gallon container", which appears to have a numeral in argument position. However, *container* is perhaps not an argument-taker here, and is being used like *bucket*, or like *magazine* in (5a). Of course then we must ask, why is *holder* not taken in the same way in (5b), and to this I have no answer.

2.3. Definite quantifiers

Definite quantifiers are quantifiers that generate a presupposition, and these are not allowed in compounds. *The* is perhaps *the* definite quantifier, but I will leave it for later and here talk about *both* vs. *two*, where *two* does not generate a presupposition. So this is what happens:

(6)   a.   Jim taught me a cast with two hands
        b.   Jim taught me a cast with both hands
        c.   Jim taught me a two hand cast[6]
        d. * Jim taught me a both hand cast

Both encodes two notions, one an existential 2 (asserting "there exists 2 N") and the other a presupposition of exhaustion (presupposing that there are no more than 2). There is a more general syntactic form for doing exactly that, namely "all n" where n is a numeral, with *both* as the dedicated way to say "all 2":

(7)   Bring me all 3 prisoners

It is the presupposition that poisons compounds:

(8)   a.   An all purpose flour
        b.   A 3 purpose flour (e.g. cakes, bread, and pasta).
        c. *An all 3 purpose flour
        d.   This is a flour for all 3 purposes

---

5   This recalls Baker's (1996) proposal that the exterior syntactic NP corresponding to an incorporated Noun is an adjunct and not an argument.
6   The conventional expression in fishing is "two handed cast". To avoid collision with this expression, I have chosen here to contrast "two hand cast" with "both hand cast", both of which are novel, but of course I regard it as non-accidental that *"both handed cast" is also ungrammatical.

Does COC work here? If it is the presupposition that is causing the problem in compounds how does COC rule it out? Essentially the presupposition has an open variable which is filled in by a) further linguistic context or b) the utterance context; in either case, within the compound it is a free variable.

Intuitively, the identification of the presupposition of "all 3 X" can vary from case to case—in some cases it is the 3 things that have been enumerated in the conversation in which the utterance occurs; in other cases, it might be the 3 generally known things. The 3 purposes of (8) are probably dependent on the conversation local to it, as I don't think that there is a universally fixed set of 3 purposes for flour (e.g. roux, children's paste, and stew thickener are 3 more purposes). But the phrase "all 3 branches of government" is not dependent on local conversational context, at least not in the US. Either way, "all 3" is context-dependent in a way that COC disallows, whereas *all* by itself, or *3* by itself, is not. I will return to these thoughts when I get to *the* and *a*.

R. Kayne (p.c.) has pointed out the following example, with *both* inside the first term of a compound:

(9) a both-hands-in game

(9) refers to a situation in a bridge game in which both players on a team are playing hands. I suspect that "both hands in" constitutes a small clause (it in fact means "both hands are in"), and that the small clause can serve as the scope of *both*, so COC is not violated. The "both hand cast" case does not have a clausal domain within the compound, and so the only way for *both* to get appropriate scope there is to violate COC. I earlier said that clauses were barred from compounds because Tense needed interpretation and COC blocked it, but "both hands in" does not have Tense and so is not barred on those grounds either.

In contrast to compounds, the [NP's N] possessive construction allows every kind of quantificational element discussed, and in all shows no COC effects:[7]

---

7 Curiously, the definite quantifier in possessive position does not allow the kind of interpretation that was discussed in connection with compounds, and likewise for universal quantifiers. Compare:

(i)   a. *All three cars' owner has disappeared
         b. The owner of all three cars has disappeared.

Apparently the possessive must distribute over the NP it is in, and that is incompatible with a singular noun head. No such restriction holds for the form in (b). A similar restriction holds for every:

(ii)   a. At the car show, I met the owner of every electric-blue car
                i. He was an eccentric character
                ii. They were eccentric characters
         b. At the car show, I met every electric-blue car's owner
                i. *He was an eccentric character
                ii. They were eccentric characters

I suspect that this is related to other respects in which a possessive does not "reconstruct", as first observed by Chomsky (1970):

(iii)   a. *John's appearance to have left (cf. John appears to have left)
         b. *John's belief [t to have left]   (cf. John was believed to have left)

(10)   a. Distributing universals are allowed:
               Everyone's favorite dessert was set at their place
        b. Definite quantifiers are allowed:
               All four players' knees had given out at the same time

2.4. Ordinals

        An ordinal (as in "the fifth winner") takes two arguments, one locally filled, and the other filled in variously—it is a relation between an individual, the first argument, and an ordered series of individuals, the second argument. So if John is the second winner, he is a winner, and there is an ordered series of winners determined explicitly (e.g. "the second of the five") or by context of some sort.   In this, it resembles the definite quantifiers we have seen in the last section, and as expected it is the second argument that interacts with COC.

Sometimes the first argument of the ordinal strongly suggest the second argument—consider "seventh inning", "sixth sense", "fifth column", "fourth estate", "third factor", "second opinion", "first responder". In all of these cases, the form [[ordinal N] N] is well-formed because the value of the second argument of the ordinal can be determined locally, and generically, and need not be postponed until local conversational context is available: "the/a seventh inning stretch", "the/a sixth sense phenomenon", "the/a fifth column subterfuge", "the/a fourth estate resurgence", "the/a third factor explanation", "the/a first responder responsibility".

But where there is no relevant convention for locally determining the value of the second argument of the ordinal, the [[ordinal N] N] compound is ill-formed:

(11)   a.   *a fifth child earache
        b.   *the fifth winner gratitude

This is because the second argument of the ordinal is free in the compound.

Compare (11) to

(12)   a. the earache of the fifth child
        b. the gratitude of the fifth winner

There is no series that "child" generically fits into, and so the second variable of the ordinal is free in (11) and (12) above; in (12) it invites speculation about what series the child or winner is a member of; in (11) it is simply ungrammatical, because of COC. "Inning" on the other hand does fit into a generic series, the innings of a generic baseball game, and so, a [[seventh inning] home run]] is fine.[8]

---

8 It is possible to dissociate a term like "inning" from its generic series, the innings in a single game. Suppose we (as sports-writers) analyzed 1000 innings of baseball play, each drawn from a different game; we could refer then to "the 241st inning" in our series, but not to the [[241st inning] home run]; but a reference to "the home run in the 241st inning" still works.

3. *More-than*

More can occur in compounds to a limited extent, but COC says that the *than* clause that goes with it cannot be linked to it from outside the compound, whether the *than* clause is overt or not.

The following is ambiguous, with the bracketing suggesting the two different meanings:

(1)   a. John is a big clock maker
        b. John is a [big clock] maker
        c. John is a big [clock maker]

However, (a) in the following is *not* ambiguous:

(2)   a. John is a bigger clock maker than I am
        b. *John is a [[bigger clock] maker] than I am
        c. John is a [bigger [clock maker]] than I am

COC predicts (b) is ruled out, because [bigger clock] contains a free variable, the argument of *more* (or rather its morphological equivalent, *-er*) that the *than* clause corresponds to. (c) on the other hand freely allows the *more-than* relation.

To justify the role of COC here, it is important to rule out other possible explanations for the failure of (b).   (b) is not bad because the embedding of *more* in it is too deep with respect to the *than* clause. First, we get the same effect if the *than* clause is simply omitted from (b)—that will still leave the *than* variable of *more* as a free variable in the first part of the compound. Second, in other examples where compounding is not involved we see that a pattern of embedding identical to the one in (b) is tolerated:

(3) John broke into a [[[more important] politician's] office] than we did.

This example also shows that the possessive construction, which we are trying to systematically contrast with compounding, continues to show no COC effects at all.

Finally, it is not simply that *more* is barred from compounds ad hocly; so long as COC is observed, it works:

(4)   a [[[more than 3] car] family]

Here the *more-than* relation complies with COC, as *more* and the *than* clause are both in the first term of the compound. There is also the issue of the scope of *more* in all these examples; I omit specific discussion here because the relation to the positioning of the *than* clause is so closely tied to the scope of *more*.

As a further point, there is a use of *more* which doesn't take a *than* clause, even a suppressed one; in "They make the bigger cars", which really means, they make the biggest cars, the *than* clause is not only optional, it is impossible. And this usage does occur in compounds: "They are one of the [[bigger car] manufacturers".

So COC makes a set of somewhat refined distinctions in this complex empirical domain.

4. Predicates with contextually bound arguments (e.g. *girlfriend, different, local*)

Girlfriend is a relational term in that x is a girlfriend of y. The y variable can be bound in several ways—by an explicit phrase, by control by an overt element, or by control by a covert element:

(1)   a. his former girlfriend
       b. Bill misses former girlfriends
       c. Former girlfriends are annoying (= One's former girlfriends are annoying to one)

There is a fourth way the y argument can be bound, by local existential closure, which we will look at a minute.

Now, the fact that y is a free variable has consequences for how the expression "former girlfriend" will behave in compounds. The only options allowed under COC is binding of the variable *from within the compound,* We will see that the restriction holds fully for compounds, but not at all for the possessive construction.

(2)   a. John met a former girlfriend killer
       b. John is a former girlfriend killer
          (only the readings where *girlfriend* is the object argument of *killer* are at issue here).

For (a), there are only two options—either the person John met killed his own girlfriend (that is, the killer's girlfriend); or, the person John met killed people who were generically former girlfriends. The first option corresponds to binding of y by the agent of killer from within the compound, and the second corresponds to local existential closure of the y argument, both allowed under COC.

What is strikingly missing is a reading in which John met someone who killed John's former girlfriend. That would correspond to binding into the compound from the matrix clause, but COC bars that.

The one "exception" to COC involves the head position of the compound—its external argument is a property of the entire compound, and so is accessible from outside of the compound. If it were not accessible the compound could not be semantically integrated with the rest of the sentence (see the parallel discussion of Tense in verb-headed compounds in section 1.3). So in (2b), the compound is predicated of *John*, which means *John* binds the agent argument of the head, *killer;* but that gives the appearance that *John* can bind the suppressed y argument of *girlfriend*; in fact though, the relation of *John* to that argument is indirect, mediated by the Agent argument of *killer*. So that is why (b) allows an apparent binding of y by *John*, whereas (a) does not.

The situation is different if "former girlfriend" occurs in a possessive phrase, rather than as the first element of a compound:

(3)   John met a [former girlfriend]'s killer

Here in fact the most prominent reading is the one where y *is* bound by *John*—exactly the reading disallowed in (2a), making the facts in (2) all the more telling about the nature of compounds.

13

Furthermore, it is very hard to get the "generic" interpretation in (3) or for that matter the interpretation in which y is bound by the agent of *killer*. The generic interpretation is generally hard to get when there is an overt potential antecedent:

(4) John met a former girlfriend

It seems that the overt antecedent wins out over other possibilities, which again makes all the more striking the availability of the generic interpretation in (2a)—it must be that the binding by overt *John* is simply not a possibility at all and so does not override the generic existential closure. And that is what we would expect from COC.

Like *girlfriend*, d*ifferent* has two arguments, and one is often suppressed— "John is different (from me)". Compounds are opaque to the binding of the *from* argument of *different*:

(5)   a. GM is a Cadillac manufacturer, but I was thinking of the manufacturer of a different car
        b. GM is a Cadillac manufacturer, but I am thinking of a
              i. *[different car] manufacturer
                    (make sure not to read it as [different [car manufacturer]])
              ii. [small car] manufacturer

This is expected under COC, and again, the [NP's N] construction is not at all opaque in this way:

(6)   I was thinking of a different car's manufacturer

Local is similar—"John went to a local bar" suppresses the argument "local to him". In a compound, this argument is inaccessible under COC:

(7)   a. John killed a local bar goer.   (only the structure [[local bar] goer] is relevant here)
        b. John is a local bar goer
        c. John met a local bar's owner

COC dictates that in (a) the bar is not understood to be local to John, but rather local to the goer. (b) evades COC by accessing the Agent argument of the head (*goer*) and so indirectly binds the suppressed argument of *local*; and (c) shows that the possessive construction is entirely free of COC effects—in fact in (c) *John* is the preferred antecedent of the suppressed argument, and I find control by the Agent of *owner* not really available in this example for reasons I do not understand.

Compounding reveals a distinction between the expressions "last year", and "the previous year". Both expressions, like Tense (see section 1.3), instantiate binary time relations, and so have two variables, one bound locally, and the other bound variously.

The difference between "last year" and "the previous year" can be seen in the following:

(8)   a. John said he won the prize last year
        b. John said he won the prize the previous year

In (a) "last year" refers to the year immediately previous to the year of the speech-time, but "the previous year" has a reading in which it refers to the year immediately previous to the event time of the matrix clause (that is, the saying time). We might say that "last year" must be "anchored" to the speech-time, meaning that its second variable must be bound to the speech-time. "The previous year" can get some other interpretations as well, including one in which it refers to a year prior to a year mentioned in discourse preceding the utterance. But "last year" always gets a single unique interpretation—the year prior to the speech-time. Similar things can be said about "last week/day/month/semester/" and "next week/day/etc" and "this week/day/etc". So, (last N) and (next N) and (this N) are always anchored to the speech-time. "The previous N" is in fact awkward in this use:

(9)   A: Are you teaching Advanced Glottochronology this semester?
        B: No, I taught it last semester / ??the previous semester

Now look what happens in compounds:

(10)   a. They are instituting a best movie of the previous year award
        b. *They are instituting a best movie of last year award
        c. They are instituting a best movie of the previous five years award
        d. *They are instituting a best movie of the last 5 years award.

In (a), "the previous year" is anchored to "award time"—awards are given in years, and one understands (a) to mean that the award will go to the best movie of the year prior to the award being given. We see immediately why (b) is bad: "last year" must be anchored to the speech-time, but since it is contained in the first term of a compound, that is impossible, as that would leave its second variable free in the compound.

A lesson of these examples is that it would be very difficult to exclude (b) syntactically, by which I mean, *constructing the rules of word and phrase formation in such a way that it was never produced in the first place*. Although "best movie of last year" is the first term in the compound in (b), it has itself a phrasal (that is, non-compound) structure, and in another environment, that structure would be fine:

(11)   a. They are instituting an award for the best movie of last year
        b. They are instituting an award for the best movie of the previous year

(a) describes an odd situation, where an award is instituted to be given only once, but it is grammatical. (This oddness follows from the required anchoring of "last year" to the speech-time.) (b) does not have the same oddness, because "the previous year" is anchored differently.

And we can add even more non-compound structure around the occurrence of "last year", but the violation persists:

(12)   a. They are instituting a [$_B$ best actor in [$_A$ a movie of *last/the previous year]] award
        b. They are instituting a
                [$_C$ best assistant to [$_B$ an actor in [$_A$ a movie of *last/the previous year]]] award

A, B and C each have purely non-compound structure, even though they are eventually made part of a compound when joined with *award*. This means that the point in the derivation when it can be

determined that "last year" is licensed by getting a proper anchoring can be as remote as you like from the point at which it enters the derivation. Which means that we probably cannot write the rules of compound formation in such a way that such things are not generated in the first place. Which means that COC must be imposed as a filter on already-generated structures. I am sure that some people will be uncomfortable with that conclusion. Maybe they can find a way around it.

Other items that have the same anchoring restriction to speech-time as "last year" and "next year" are "5 days ago", *now* (but not *then*), *here* (but not *there*), and *yesterday*, *today* and *tomorrow*. *There* and *then* have more anchoring possibilities than *here* and *now,* but since they are pronouns they won't do well in compounds either, unless you could construct a compound big enough to contain an antecedent for them. Compare:

(13)   a.   The June 5th eruption
       b. *The yesterday/today eruption
       c.   The New Jersey eruption
       d. *The here eruption
       e.   The New Jersey June 5th eruption
       f. *The here yesterday eruption

By contrast, the possessive construction shows none of the COC effects we have demonstrated for compounds:

(14)   a. *local* can be bound from outside:
          John met a local bar's owner (local to John)
       b. *girlfriend* can be bound from outside:
          John met a former girlfriend's murderer (John's former girlfriend)
       c. *different* can be bound from outside:
          I was thinking of a different car's manufacturer
       d. *Next* and *last* can receive speech-time anchoring:
          Next year's heroes will be the children

5. Reflexives

Of course reflexive pronouns, which are dedicated variables, cannot occur in compounds. *Or can they???!!!*   It turns out that there is a reflexive form in English that does occur in compounds, and it is instructive to see why it is possible[9]:

(1)   a. self-education
       b. John believes that self-education is the only education.
       c. John's self-education = John's education of himself
       d. John likes self-education stories
       e. John likes stories about the education of himself

9   See D&W pp. 58-59 for more discussion of the difference between *himself* and *self-*

First we see that *self-* obeys COC, as it it is bound within the compound within which it occurs. In (a), *self-* is bound to the Agent argument of *education*, and itself satisfies the Theme argument of *education*. (b) shows that that compound-internal binding is sufficient even when the Agent of *education* is suppressed. Maybe it would be more proper to say that *self-* induces the binding of one of the arguments of *education* to another of its arguments; either way, COC is satisfied. Even on the first view, there is no term of a compound that *contains* a free variable; even though *self-* itself is a free variable.

Example (c) shows that *self-* can be linked to an overt expression, *John* here, but of course this would be via the indirect link through the external argument of *education*—we saw this before in the discussion of "local bar" in the previous section.

Examples (d) and (e) show a striking difference between *-self* and *himself*—only *himself* can be bound to overt antecedent *John* without routing through the Agent argument of *education*. Only (e) can mean "stories about someone else's education of John". To get that interpretation, *himself* must remain a free variable in both "education of himself" and in the larger NP, "stories about...", not being bound until the matrix clause is reached. COC guarantees that no such possibility exists for *self-*.

There is a remaining mystery though; COC would not prevent the full reflexive *himself* from appearing in a compound, so long as it was bound to an argument of the head, as in;

(2) *John's himself education

Here *himself* is identified with the Theme argument of *education*, and bound to the Agent argument of *education*, and ultimately indirectly bound to *John* via the identification of *John* with the Agent argument. Everything works, and complies with COC, so (2) is unexplained. A possible route to understanding this arising from Kayne (2002) is that what needs an antecedent in the case of *himself* is not *himself* as a whole, but rather the *him* within it. In that case *himself* would be an invalid term of a compound, containing as it does the free variable *him*. This doesn't square everything though, because we now have *himself* as a compound with a free variable in it, and COC should block that as well, just as it blocks "it robber". So finally (2) remains unexplained. It may be that *-self* is an not an anaphor but rather an operation of "argument reduction" on the word that it attaches to, whereas *himself* requires an overt syntactic antecedent.

By contrast with compounds, the possessive construction is systematically transparent to anaphor binding. The English reflexives cannot appear in the possessive position, but this is likely due to a case restriction—the reflexive appears only in accusative positions, and so is excluded in nominative positions as well (*"John thinks that himself won"). But the reciprocal expression "each other" is not so restricted, and does occur as a possessive:

(3)   They admired each others' wardrobes.

6. Articles in compounds.

The most strikingly unbreakable restriction on compound formation is that neither the right nor the left term can have an article. By "article" I mean *a* and *the*.

17

(1)   *the [[a building] smasher]
         *the [book [the printer]]

However, it is possible to have an articled NP inside either of the terms of the compound:

(2)   a. a [[sale of [the day]] award]
         b. a [New York City [man of [the hour]]]

It is tempting, especially given what has preceded, to say that scope is the culprit—these articled NPs must take scope; the compound itself is not a valid scope domain, and assigning it any higher scope would violate COC; but the left or right term of the compound could serve as a scope domain for something within them without violating COC.

However, an R. Kayne (pc) example raises a doubt about this otherwise plausible line—even in an expression like "The Bronx", the article is disallowed in the first term of a compound:

(3)   a. *The [[The Bronx] zoo]
         b.   The [Bronx zoo]

It seems unlikely that "The Bronx" would take scope, when Brooklyn, for example, would not. In the case of The Hague, even omitting the article seems impossible, leaving no way to compose the expression:

(4)   a. *A [[The Hague] meeting]
         b. *A [Hague meeting]
         c.   A [Pittsburgh meeting]

Using scope to account for these differences starts to look doubtful.   Nevertheless, I will put both The Bronx and The Hague aside and pursue the scope account to take things one step in what I hope is the forward direction.

        Returning to examples like (2), we notice that the articled NPs that do show up inside compound terms are sharply limited in their interpretation. When we hear an expression like the following:

(5) We are going to establish a [[movie of the year] award]

We don't ask "which year"— we understand "the year" only in a strictly generic sense.   But in a general syntactic context we find genuine ambiguity for definite time expressions:

(6) I thought John would leave by the end of the year.

In one (kind of) interpretation of (6), "the year" is bound to the matrix Tense, "anchored" to it in the sense of section 4, and so it means the end of the year in which my thought occurred. But in another context, "the year" could refer to a particular year established by the context of the conversation, a year under discussion:

(7) a. (In 1966) I knew that John was coming to work for us in 1967,
             but I thought that he would leave by the end of the year (that is, the end of 1967).
     b. I knew that John would come to work for us sometime in the 60s,
             but I thought that he would leave by the end of the year
             (that is, the end of whatever year it was that he came)

We might attribute the different interpretations of "by the end of the year" in (8) to scope—if the definite time expression has more local scope, it is interpreted with reference to a Tense operator; but if it takes widest scope, its referent can be resolved by properties of the utterance context.   In any case, "the year" has a free variable that needs to be resolved in larger context.

It is of course well beyond the ambition of this note to assess the viability of such a line of thinking. It is nevertheless relevant to present purposes that neither of the sorts of interpretation (Tense-bound or utterance-bound) that was found for (6) and (7) is available for an articled NP within a compound term. The articled NP must get scope, but thanks to COC the scope is so local that neither c-commanding Tense nor the utterance context can be involved in its interpretation (that is, to anchor it).   The interpretation available is similar to what "the day" gets in fixed expressions like "At the end of the day" and "back in the day"—no one asks, "which day?".

So, COC prevents an articled NP from being either term of a compound, and it sets a sharp limit on the interpretation of articled NPs that occur within a term of a compound by preventing them from interacting with the linguistic and utterance context in which the compound occurs, leaving them with at most generic interpretations.

7. Why does COC exist, and why does it apply to what it applies to?

Throughout I have tried to show, whenever feasible, that COC effects found with compounds were entirely absent from the possessive construction.   To enumerate them all in one place, compounding, but not the possessive construction, is opaque to:

(1) a. definite quantifiers (*both, all 3,* ordinals)
     b. distributing quantifiers
     c. *more...than*
     d. anaphor binding
     e. the binding of suppressed arguments of *girlfriend, local, different*, etc.,
     f. other indexicals like *now, then, next, today and tomorrow*
     g. pronouns
     h. the shifting sense of "Miss America"
     i. numerical quantifiers in argument position
     j. articled NPs
     k. Tense

This systematic contrast with the possessive construction sets limits on how COC could be re-conceived or replaced. For example, it is not possible that it is the nominality of compounds that is responsible for COC effects—first, not all compounds are nominal, but more importantly the possessive construction is just as nominal, and shows none of the effects. And COC effects cannot be somehow attributed to the fact that the compound instantiates a wide range of linguistic relations

19

between its terms—argument-of, adjunct-of, and others less nameable; the possessive construction famously instantiates nearly as wide a range and shows no COC effects.

This is not to say that there are no "generic" instances of the possessive construction—Liberman and Sproat (1992) is full of them: *pope's nose, servant's entrance, cook's tour,* etc. In addition, English had a now defunct Romance-style source of generics terms: *scoff law, dare devil, cut throat.* But the right-headed compound construction in today's English seems to have been dedicated to serving up generic terms, and COC is what enforces it.

But this raises the question, why should compounding submit in such a thorough way to COC? Wouldn't language be simpler, and at the same time capable of more, if COC didn't exist?

I have talked about COC as inducing genericity in compounds, as though genericity were a by-product, but now I think the reverse perspective is worth exploring. Maybe genericity is the goal, and COC is merely the means.

Language has a stock of generic terms (*dog, on, give, sad*) whose interpretations are "assignment independent". It looks like a functioning language needs a few thousand of these.

It also looks like English has devoted the [N N] structure (and more generally [X head]) to producing novel generic terms. D&W thought that there was a close connection between head finality and genericity, and from that arose their version of the "lexicalist hypothesis". But that might have been hasty. First, even just within the realm of English compounds, we find terms that have head-initial syntax, but have enforced genericity anyway (e.g. "more than 3 car family", and many other examples discussed above).

In addition, Romance does not have head-final compounding, but instead constructs generic expressions from its phrasal syntax ("machine a laver", "machine to wash", French for washing machine) It may be that French has no special syntax at all for generic terms and uses exclusively phrasal syntax for new generic terms; the limitations on them may all follow from COC.

Maybe every language has a way to generate new generic terms, but different languages deploy different resources that are already available. The English compound system is a perhaps an extension of the English prefixation system, with the non-head of the compound serving as the prefix. The French system is instead a contraction of the phrasal syntax system (contraction because anything that violates COC will be filtered out). COC then applies to both, because it is tied to the universal goal of providing generic vocabulary, not to the language particular means.

D&W thought that a certain kind of genericity and a certain kind of syntax—in English, head final syntax—came together in an inevitable package ("words"). But I now see this as mistaken. In this note I have restricted myself to compounds, but have not said what a compound is. The usual definition is that it is a right-headed binary structure that has either words or other right-headed binary structures as its members; the following are typical:

(2)   a. [dog$_N$ park$_N$]
       b. [[dog$_N$ park$_N$] [bathroom$_N$ facility$_N$]]$_N$
       c. [[[dog$_N$ park$_N$] [bathroom$_N$ facility$_N$]$_N$]$_N$ [utility$_N$ brush$_N$]$_N$]$_N$

These conform to the following simple rule (a); but for generality in discussion, let's take (b):

(3) a. N → [N N]$_N$
   b. X → [Y X]$_X$ where X and Y are lexical categories

If this is all you have got, it is conceivable that COC could follow from the impoverished syntax of these expressions; quantifiers and operators like *-er* and so forth don't come into the picture because of the minimal syntax. But in this note I have reviewed a number of compounds that are not restricted in the way (3b) would suggest; here is a summary listing:

(4) a. The Miss America Show
   b. an all purpose flour
   c. a 6 drawer chest
   d. a both-hands-in game
   e. a [[[more than 3] car] family]
   f. John met a former girlfriend killer
   g. a [small car] manufacturer
   h. John is a local bar goer
   i. They are instituting a best movie of the previous year award
   j. They are instituting a [$_B$ best actor in [$_A$ a movie of the *last/previous year]] award
   k. a [[sale of [the day]] award]
   l. [New York City [man of [the hour]]]
   m. the [man of [the hour]] award]

Take (l)—we want to call it a compound because of the opacity that was demonstrated for it, and because it is right-headed. But it has a left-headed expression in its head position. In (k), that same left-headed expression appears as the left member of a compound. The generality with which forms that do not conform to (3b) can occur as members of a compound term begin to make the rules in (3) seem useless; especially the restriction of X and Y to lexical categories. What we really need is:

(5) X → [YP XP]

where XP and YP can be any expressions in English at all; what is distinctive about (5) is only its right-headedness plus the determination that the result X is a lexical category, meaning that it can occupy the lowest head position of XP.

If we adopt (5) as the rule for compounds, then we have given up on deriving COC from the internal structure of compounds. (5) will highly over-generate, but we cannot cut back on it by revising (5), because we still need to derive every form in (4). But we already know that COC can do most if not all of the cutting back we need. But this means that COC cannot be derived as a consequence of any aspect of the structure on the right-hand side of (5).

The one thing about (5) that we might hang the COC on is the left-hand side; that is, the fact that the structure on the right composes a member of a lexical category. We might regard COC then as a condition on lexical categories which has nothing to do with the syntax on the right-hand side. If this is the case then have an explanation of the observation at the end of section 4 (concerning example (4.12),

repeated as (4j) above) that an illicit free variable can be be buried deep inside non-compound-looking structure and nevertheless runs afoul of COC if it has not been bound by time the derivation reaches the level where the first right-headed compound structure is put together.   That is spooky action at a distance.

Rule (5) plus the application of COC to it is exactly the strategy D&W used to treat French compounds (of the sort *coup-de-grace (NP), essuie-glace (VP)* "windshield wiper")[10] ; there the proposed rule was:

(6)   N $\rightarrow$ XP   (D&W p. 84)

and, as here, the over-generation was cut back by the atomicity condition, and for the same reason: the left-hand side is a lexical category.   What is changed here is that English compounds look more like French compounds, in having access to the full phrasal repertoire.[11]

It this light, we should now generalize the COC to the OC (Opacity Condition), a condition on X^0s, bringing it fully in line with the D&W atomicity condition cited in (0.1) above.

Languages do not need dedicated means to produce generic expressions. Any expression that happens to satisfy OC will do. For example, "play pianos" (as in "They play pianos") satisfies OC and therefore qualifies as a generic expression, but no special syntax is required, as the bare plural is independently available ("Pianos are versatile"). For other forms, it is less obvious; it might be thought, for example, that "play piano", as it occurs in (a) in the following, is licensed as a special dedicated generic term:

(7)   a.   John plays piano
        b.   John plays pianos
        c. *John buys piano
        d.   John buys pianos

But in fact (a) needs no special license: the only thing special about (a) is that *piano* is being used as a mass term, as can be seen clearly in the following:

(8)   a. John plays lots of piano these days
        b. John plays more piano than guitar
        c. With Gershwin's music, the more piano the less romantic.

(8c) shows that *piano* is licensed generally as a mass term independent of *play*, so (7a) needs no special story—if anything does it is (7c). But in fact, in the right context, (7c) is acceptable; if John is a sheet music collector (c) is fine but (d) is not. The special thing in all this is that the mass noun *piano* refers to a type of music, not to a type of instrument.

10 D&W did not notice that there are fixed expressions in this class containing proper names, such as "violon d'Ingres".
11 Actually D&W propose (6) for English as well, but for expressions like "breakdown" and "push up"

I conclude therefore that "play piano" is simply a syntactic phrase, and not, for example, an instance of "quasi incorporation" as has been proposed for "spelen piano" in Dutch (Booij 2008), and it appears that no special syntax is needed for it.

So, what the rules in (5) and (6) provide is not generic expressions, which are going to be available anyway, but rather a source of generic expressions that can be used as lexical heads, or X^0s. This is where English differs from French—English has enlisted a right-headed construction, $[XP\ YP]_Y$, as a source of X^0s, where that construction is not a part of its phrasal inventory. Maybe this reduces to some other parametric difference between English and French, maybe not.

To close this section, I will briefly review three additional phrasal sources of generic terms in English. Only the first of them gets a satisfactory account under OC.

In fact the possessive construction is the first of them, in spite of my having used it to systematically contrast with $[N\ N]_N$ compounds. As mentioned earlier, Liberman and Sproat (1992) cite a large number of fixed expressions having [NP's N] form:

(9)   a. pope's nose
      b. servant's entrance
      c. bishop's mitre
      d. $N \rightarrow NP$'s N

As the rule in (d) suggests, these possessives can be used a though they were simple nouns inside a larger NP with a second possessor:

(10)   Bill's huge pope's nose/ bishop's mitre/servant's entrance

As indicated in (d), the special stipulation for English is that its possessive construction can be used to construct new members of the Noun lexical category. We may assume that OC will apply automatically, if if it is a universal property of lexical categories.

These possessives are not a frozen set of expressions; it is productively extendable, if not quite to the wild extent that N-N compounds are. For example, I have never heard of a hospital having a "doctors' mail room", and doubt if there is such a thing, but I can straightforwardly say,

(11)   The new hospital has a huge new doctors' mail room
       That hospital's huge new doctors' mail room

The new [NP's N] form is being used as the head noun of an NP with its own possessive. Clearly, the productive phrasal "NP's N" has been specialized to productively produce new generic terms that serve as Nouns. The specialization is effortless, as OC does the main work. Despite being a productive category, these possessives will have none of the privileges of regular possessive constructions, because of OC. For example,

(12)   a. The nurses are envious of the huge new doctors' mail room
       b. The nurses and the doctors are envious of each other's huge new mail rooms
       c. *The nurses and the doctors are envious of the huge new each other's mail rooms

d. The doctors are happy about their huge new mail room
e. *The doctors are happy about the huge new their mail room

As a second case, I suspect that [A N] is another source of generic terms. If [A N] were exclusively generic in the sense that OC defines, it would explain why when operators with complements, like *too* and *so,* attached to an adjective, a special syntax is required:

(13)   a. *John is [a [[too honest] person] to lie][12]
        b.   John is [[too honest] [a person] to lie]

But there are hurdles looming immediately: similar operators don't trigger the special syntax:

(14)   a.   John is an honest enough person not to lie
        b.   John is a more honest person than I thought
        c. ??John is more honest a person than I thought

But it is at least suggestive.

And finally, there is the pseudo-passive construction, already identified as a candidate in D&W, stemming from proposals of Hornstein and Weinberg (1981) and Chomsky (1973) that "the reanalyzed material of a passive must be a "natural predicate" in some sense" (p. 54); the examples cited are these, with the reanalyzed material in brackets:

(15)   a. Bill was [yelled at] t
        b.*Bill was [sent some books to] t
        c. Bill was [taken advantage of] t

We should add "natural predicate" to the list of terms that includes "(natural) kind", "characteristic", "nameworthy", and "generic".   D&W suggested that the needed notion of "natural predicate" was captured by their atomicity condition, which I have argued OC to be an improvement on.

Unfortunately though, OC is in a much weaker position than the D&W atomicity condition, especially in light of the findings of this note. First, it is not clear that "some books" in (b) could not take scope within the bracketed material, which would render it compliant with OC, but not with the D&W atomicity condition, because that condition said that no scope assignment at all was admissible for items in the domain in question, even very local scope. Second, we have seen earlier that proper names, unlike pronouns, do not give rise to OC violations, and yet the following is at least as ungrammatical as (b):

(16)   *Bill was [recommended Sally to] t

In regard to the pseudo-passive then, it looks like I am worse off now than in 1987. So we beat on, theories against the facts, borne back ceaselessly into the past.

---

12 R. Kayne (pc) points out that "a much too honest person to lie" is grammatical.

Bibliography

Anderson S. 1992 *Amorphous Morphology* Cambridge University Press

Baker, M. 1996 *The Polysynthesis Parameter*. New York: Oxford University Press.

Barrie, M. and E. Mathieu 2016 "Noun Incorporation and Phrasal Movement" in NLLT 34:1-55

Booij, G. 2008 "Pseudo-incorporation in Dutch." *GAGL: Groninger Arbeiten zur germanistischen Linguistik*, 46: 3–26.

Chierchia, G. 1998 "Reference to Kinds across Languages" *Natural Language Semantics* 6: 339–405

Chomsky, N. 1970 "Remarks on Nominalization" in R. Jacobs and P. Rosenbaum *Readings in English Transformational Grammar* Ginn and Co.

Chomsky, N. 1973 "Conditions on Transformations." In S. Anderson and P. Kiparsky, eds., A *Festschrift for Morris Halle*. Holt, Rinehart and Winston, New York.

Déchaine, R-M. & M. Wiltschko. 2002 "Decomposing Pronouns" *Linguistic Inquiry* 33:409-422.

Di Sciullo, A-M. and E. Williams 1987 *On the Definition of 'Word'* MIT Press

Dowty, D., R. Wall, and S. Peters 1981 *Introduction to Montague Semantics* Springer

Goddard, I. 1990 "Primary and secondary stem derivation in Algonquian" *International Journal of American Linguistics* 56: 449–483

Haegeman, L and H. van Riemsdijk 1986 "Verb projection raising, scope, and the typology of rules" *Linguistic Inquiry* 17: 417–466.

Hornstein, N., and A. Weinberg 1981 "Case Theory and Preposition Stranding" *Linguistic Inquiry* 12, 55-91.

Johns, A. 2007 "Restricting Noun Incorporation: Root Movement." *Natural Language and Linguistic Theory*, 25 (3): 535–576.

Kayne, R. 2002 "Pronouns and Their Antecedents" in in S. Epstein and Daniel Seely, eds, *Derivation and Explanation in the Minimalist Program* Wiley, pp 133-166.

Kripke, S. 1980 *Naming and Necessity* Harvard University Press

Liberman, M and R. Sproat 1992 "The Stress and Structure of Modified Noun Phrases in English" in *Lexical Matters*, I. Sag and A. Szabolcsi, eds., Stanford University.

Mithun, M. 1984 "The Evolution of Noun Incorporation" *Language*, 60: 847–893

Reichenbach, H. 1947 *Elements of Symbolic Logic* Macmillan

Sadock, J. 1980 "Noun incorporation in Greenlandic: A case of syntactic word formation" *Language* 56: 300–319.

Sapir, E. 1911 "The Problem of Noun Incorporation in American Languages" *American Anthropologist* 13: 250–282.