A Paradox of English Determination: the Construction of Complex Number Expressions

Michael Ellsworth, Russell Lee-Goldman, Russell Rhodes
University of California, Berkeley

1. Introduction

In this paper, we investigate some points of interaction among complex number constructions. Possibly *the* most thorny issue, for English, is the question of determiner agreement. In simple cases, determiners (like *a, this, those*) agree with the number morphologically encoded on their noun. But in English (as in many other languages), numbers interact with agreement in non-straightforward ways. An example can be seen in (1a), in which multiple constructions interact to produce a singular-agreement determiner with a plural noun. As can be seen by comparison to other similar phrases, neither the singular agreement of the determiner *a* (1b), nor the grammaticality of range expressions combined with any kind of shared determiner (1c) can be taken for granted.

(1) a. an absolutely mind-boggling million to billion star clusters
b. these/*a thirty to forty clusters
c. *the/*those/*a (from) million to billion clusters

Before we can discuss how the constructions interact to produce the attested grammaticality (section 5), we must first describe the constituent constructions: complex number expressions (section 2), the *whopping*-pattern (section 3), and the range construction (section 4).

2. Number expressions

There are several different basic classes of numbers: [1]

- Simple numbers: *one, seven, eighty-seven* (all numbers < 100, excluding *dozen*.)
- Dependent numbers: *dozen, hundred, thousand, million*
- Complex number expressions: *one million, four hundred, three thousand*

forty-five; complex numbers have a multiplier (***four*** *hundred*) and may have an addition (*four hundred* ***(and) fifty***); see (6).

Simple numbers and complex numbers have identical external syntax; both, for example, require and project nominal structures which, as usual, have number agreement between determiner and head noun (2a) and allow "N' ellipsis" (2b), i.e., they can be pumped to NP (see Michaelis (to appear) on a similar type of pumping construction). They are also both usable as multipliers in the Complex number construction (2c).[2]

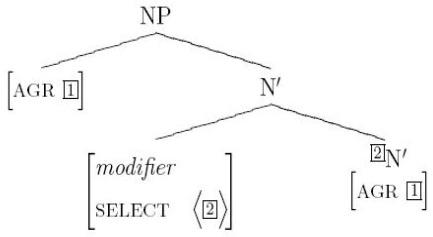
(2) a. those/#that **three (hundred)** people/*person
b. There were **three (hundred)**.
c. **three (hundred)** million

The external syntax of Dependent numbers is somewhat different. There are essentially two patterns: 1) they can be multiplied, creating complex numbers (3a), or 2) they can combine with an N' (as the other numbers do), but in this case they produce an N' that requires a determiner (3b)—most surprisingly, when the NP is indefinite, they require the singular determiner *a* (3c). As always, determiners may not iterate (3d).

(3) a. Three thousand admirals participated.
b. The thousand admirals who participated were....
c. A thousand admirals participated.
d. *The a thousand / A the thousand...

We capture the behavior of numbers with the following constructions (4-6). In (4) we show the usual case wherein a nominal modifier selects an N', producing an N' with identical CAT features (including part-of-speech and morphological marking); the specifier in the leftmost branch, furthermore, selects the upper N' whose agreement characteristics are identical to those of the head noun.[3]

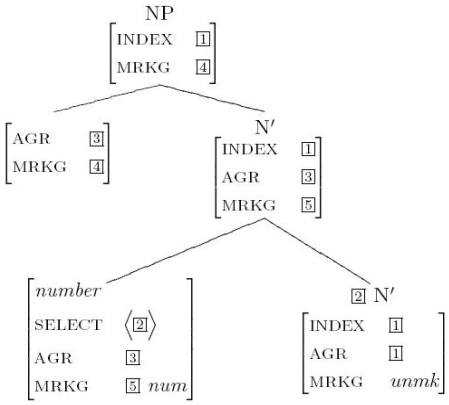
(4) Normal Modifier with N' and specifier



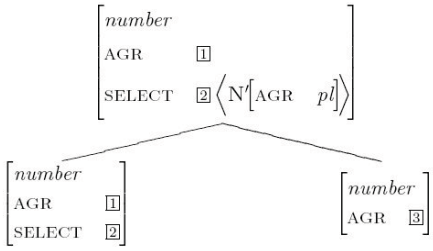
In (5), we describe how numbers combine successively with a head noun and then with a determiner whose agreement selection may not match that of the

original noun. The basic simplifying assumption is that the difference between the simple and dependent numbers is essentially what kind of determiner agreement their lexical entries require. The most important point to note is the AGR feature of the specifier (the determiner, for all purposes herein), which must be identical to that of the number, not the AGR value of the noun, as in (3c). This allows us to write lexical entries for dependent numbers (e.g. *hundred*) which idiosyncratically have singular agreement features, while the simple numbers (e.g. *twenty*) have agreement identical to the N' they select. Verbal number agreement instead uses the value of INDEX, which is the same on the original noun and the resulting NP (hence plural verbal agreement for *a hundred people know*).[4] The marking (MRKG) *num* ensures that one cannot iterate application of numbers (*the *three five* pianists).

(5) Numbers with N' and specifier



(6) Complex number construction



In (6), we show the construction that builds complex numbers, taking its AGR value from its left daughter (note the difference between *a hundred thousand*

and **a two thousand*). In combination with (5) and the AGR values of the lexical entries for simple and dependent numbers, the patterns of singular and plural agreement exemplified in (2) and (3) are explained.[5]

3. The *whopping* pattern

Consider now the sentences in (7), which seem to violate the constraint that *a* is paired with a dependent number.

(7) a. an [amazing/outstanding/whopping] [six/sixty/six hundred/hundred] admirals
b. [det] [AdjP] [number expression] [nominal]
c. a [tall/smart/loud] [six/sixty/six hundred/hundred] kids

(8) a whopping one/*a thousand admirals

With an intervening adjective like *whopping*, the NP may have a singular determiner such as *a* regardless of the number of the head noun. More strikingly, any variety of number expression—simple, complex, dependent—becomes possible, except for determined number expressions (8), which we have analyzed above (5-6) not as types of number, but rather as full NPs. Note that the ungrammaticality of (8) provides further evidence that despite any functional similarity between *a* and *one*, they are syntactically distinct.[6]

We do not here explore the semantic restrictions on the sort of adjective that may participate in this construction. There do seem to be several classes, outlined in (9), but the analysis remains in the early stages. Most (9a-d), but not all (9e-f), describe the quantity itself.

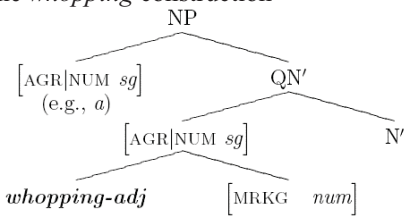
(9) a. Quantity increment: additional, further, extra
b. Quantity description: mere, scant, paltry, good, full, whole, generous
c. Typicality: unprecedented, estimated, typical, reported
d. Affective: whopping, amazing, outstanding
e. Event duration description: quick, busy, hectic
f. Description (of item) dependent on quantity?: a lucky three
g. *a tall/short/intelligent thirty students

Among the morphosyntactic properties of this construction that must be described is the fact that, with an intervening adjective, the NP may have *a* as a determiner, regardless of the number of the head noun or number. This further confirms what the previous section showed: not only is NP-internal agreement not purely based on the agreement features of the determiner and head noun, but in some cases, like those in (7), even the inherent agreement features on a number may be overridden.

The construction must further allow for any variety of number expression: simple, dependent, or complex, with the exception of determined number

expressions (8).[7] We capture all the above facts with a special class of adjectives, which we here dub *whopping-adjectives* (meant to include, perhaps with subclasses, all those varieties outlined in (9)), illustrated in (10).

(10) The *whopping* construction



The *whopping-adj* specifies its own agreement value (singular), and whatever specifier the number or head noun might have expected is "swallowed up." Then, by virtue of the construction in (5), or by other QNP-producing constructions in the language, the specifier is forced to show singular agreement. Note that because the adjective selects a *num*-marked, and not a *det*-marked, constituent, *whopping* patterns with determined numbers are ruled out (11).

(11) a whopping {one/*a} thousand admirals.

We see from the constructions set out in this and the previous section that the number features that must hold of the NP's determiner is a function of exactly which number expression constructions, including the *whopping* pattern, license the NP. The picture is considerably more complex than even that described in Kim 2003, in which NP-internal agreement is mediated by a single feature that remains constant throughout the NP. Further exploration of this topic should include the integration our NP-internal findings with the facts of NP-external (e.g., subject-verb) agreement (on which Kim 2003 provides much of the requisite data and analysis).

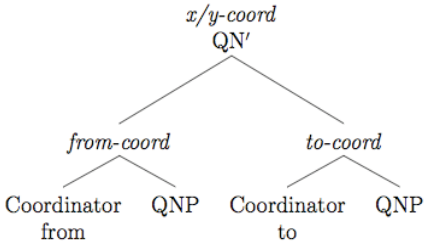
4. Range-denoting *from-x-to-y*

Consider now the range denoting *from-x-to-y* construction, which is illustrated in (12). In general, any place where *x* can go, this larger construction can go as well.

(12) a. Generally, (from) 30 to 40 people show up.
b. You might see (from) ten admirals to twenty admirals every night.
c. Just a quote will run you (from) one (hundred) to two hundred dollars.
d. He ran (from) a dozen to a hundred miles a week.

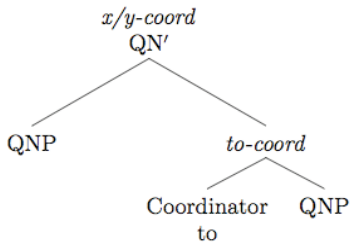
As argued by Hirose (2007), this is a correlative coordinate structure, comparable to *both...and* and *neither...nor*[8], as shown in (13). (Here and below, QN' is a convenient abbreviation for an N' with MRKG *num*, and analogously QNP is an NP built from such an N'. See (5).)

(13) The *from-x-to-y* construction



Note that there is another version of this construction without *from*, also illustrated in (12). The structure of this version is given in (14).

(14) The *x-to-y* construction



As seen in (13) and (14), the mother node in this construction is a QN'. This accounts for three distributional facts: 1) the combination with a determiner (15a), 2) *from-x-to-y* constructs cannot have a multiplier (15b-c) because the complex number construction, which introduces multipliers, selects a number rather than a QN', and 3) the construct can be pumped to a determinerless NP (12).

(15) a. the/those (from) three to four people
b. *a/five from hundred to thousand dollars
c. *a/five from dozen to a/five hundred dollars

The fact that the coordinates are QNPs accounts for several restrictions on *from-x-to-y*. Dependent numbers, as they are not QNPs, cannot stand alone in the x/y slot, as illustrated in (16a). (16b) shows that an alternative analysis in

which the coordinates are simply numbers is untenable, as it would predict that a single element can satisfy a constraint from both coordinates—in our case, a lone multiplier applying to two dependent numerals (cf. the grammatical (16c) which shows a single determiner applying to coordinated count nouns).

(16) a. *from hundred to thousand star clusters
b. *a/five hundred to thousand star clusters
c. my friend and neighbor

5. *Whopping* plus *from-x-to-y*

Having analyzed two rather different number expression-licensing constructions, we turn now to the possibility of combining them: that is, determining if a single number expression may be licensed by both constructions simultaneously. This is, at first blush, theoretically possible: the sister of a *whopping-adj* must have a marking value of *num*—i.e., it must be a number expression. This is exactly the sort of expression that is produced by the *from-x-to-y* and *x-to-y* range constructions.

Placing a *(from)-x-to-y* expression into a *whopping* expression, however, results in a grammaticality pattern that is not entirely predictable:

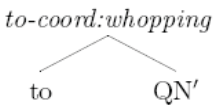
(17) a. * a whopping from 3/30/300 to 4/40/400 admirals
b. a whopping 3/30/300 to 4/40/400 admirals
c. a mind-boggling million to billion star clusters
c'. * a million to billion star clusters
d. a whopping dozen to two dozen admirals
d'. * a dozen to two dozen admirals
e. * a whopping a hundred to a hundred fifty balloons

Several restrictions are placed on the sister to the *whopping-adj* that would not have been predicted if the two constructions simply combined by general means. First, as illustrated by (17a,b), the *from*-less range construction must be used.[9] Second, although dependent numbers are normally prohibited in *x*/y slots of the range construction (16a,b), they are permitted with the addition of a *whopping-adjective* (17c/c', 17d/d'). Finally, we see that determined numerals are no longer permitted (17e), though the range construction normally allows them (12d).

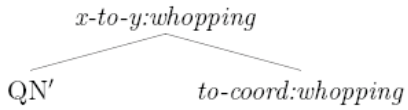
The existence of idiosyncratic constraints on the combination of these two particular constructions means that a separate set of constructions must be posited. In particular, two constructions are necessary. First, there is a special lexical entry for the coordinator *to* that selects a QN' (a *num*-marked non-maximal (undetermined) nominal expression, such as *three hundred* or simply *hundred*). Call this lexical entry *to-coord:whopping*, a subtype of *to:coord* (shown in (18a) post-combination with a QN'). It produces phrases such as *to two dozen* or *to million*. The phrase projected by this lexical entry will only be

licensed as a right daughter of another construction (18b) that combines the *to-coord:whopping* phrase with another QN', which appears to the left. The result, something like *dozen to two dozen*, is analogous to the *x-to-y* construction in (14), but with QN' rather than QNP coordinates. This structure is further restricted to only appear as the sister of a *whopping-adj*, the result of this indicated in (18c).[10] This, along with constraints discussed in previous sections, guarantees the pattern of grammaticality illustrated in (17).

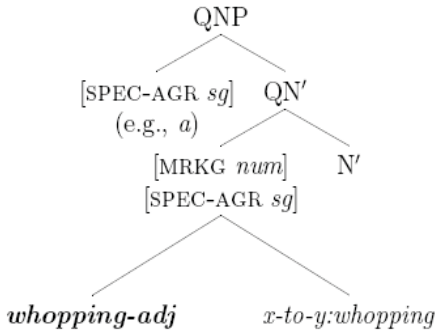
(18) a.



b.



c.



6. Conclusions and Outlook

Whether one comes from a generative or construction grammar tradition, grammar is always considered compositional, in the sense that sentences should be explainable by a distinct set of grammatical processes and/or interrelations. However, it is unpredictable what level of detail is necessary to explain patterns of interpretation and grammaticality.

In such cases as the one investigated here, it is tempting to hypothesize that the constructions we identified in sections 3 and 4 (the *whopping* construction and the *range* construction) are the *real* constructions, and the facts in section 5 are somehow derivative—i.e. due to some principle beyond the composition

described in the constructions themselves. There is, however, no way to predict *a priori* that, instead of just being incompatible, there would be a special version of the *range* construction (18) that would be usable only with the *whopping* construction, structured specifically to resolve the conflict. If incompatibilities were freely resolvable in this way, there would be no ungrammatical sentences. Although there is an interesting relationship (both semantic and structural) between the regular *range* construction and the *x-to-y:whopping* pattern, the full interpretation of how and what kind of conflict-resolving special constructions can be created lies outside our scope here, in the realm of historical linguistics and acquisition. Since limitations on the kinds of historical relationships that can arise are not part of syntax, it is important for our syntactic theory to be agnostic about how detailed our constructions will have to be, and be capable of representing whatever level of detail is necessary to model the data.

It is clear that further complexity will be necessary to explain even the narrow topic that we focus on here, the morphosyntax of NP-internal agreement. Within the constructions we posit, the restrictions on the types of adjectives that fit the *whopping* pattern and their semantic contribution to the noun phrase must be fully spelled out. In addition, our account must mesh with subject-predicate agreement, which is affected by both syntactic and semantic features, and with the wide range of NP-internal facts summarized by Kim (2003). Only in a theory unencumbered by stipulations of possible complexity will such a full description of agreement be possible.

7. Notes

[1]For the basic categories involved, we largely follow Huddleston and Pullum's (2002) 19.5.10, other than grouping together *simple*, *derivative*, and *compound* together as *simple* (since these distinctions are morphologically rather than syntactically relevant). Our analysis of number composition is different, however. In particular, Huddleston and Pullum's (2002) assumes that *a* is equivalent to a multiplier for building complex number expressions from dependent numbers. Our data show that this is not so simple. See section 3 below.

[2]Clearly considerably more detail must be injected to restrict which complex numbers can actually be produced (e.g. the multiplier must be smaller than the base in complex numbers). These details are irrelevant here.

[3]Although in (4-5) we have represented an AGR feature on the specifier itself, this is purely for convenience of presentation. Our analysis is entirely compatible with proposals in HPSG (Van Eynde 2006) and the current SBCG assumptions (Sag 2007) in which the specifier has a SELECT feature which selects an N' with the specified agreement value.

[4]The implementation of the constraints on AGR in SBCG require the *hd-func-cxt* (Sag 2007: 54) to constrain AGR to be identical in the *functor* daughter and the phrasal mother node. Virtually all other constructions would constrain AGR to be the same in the *head* daughter and the mother node. There is no structural difference between this and the behavior of the MRKG feature. An alternative requiring no change in the constructions would be to make this a subfeature of MRKG, which would then take non-atomic values.

[5] There is, in point of fact, a complication in that, for some speakers, simple numbers may also take singular agreement specifically when they occur in definite noun phrases: *that hundred senators, #that thirty senators, *a thirty senators*. This fact does not require further constructional gymnastics. With the constructions we already have, an additional simple-number lexical entry with AGR *sing*

can encode this constraint by SELECTing an N' which is definite.
[6]See Huddleston and Pullum (2002: 1718) for the more usual analysis in which *a* is considered a mere variant of *one*.
[7]This constitutes further evidence that despite any functional similarity between *a* and *one*, they are syntactically distinct.
[8]We follow Huddleston and Pullum's (2002) phrase structure for correlative coordinate structures. On their account, the *from* should be a determiner. However, because the placement of, e.g., *both* is slightly broader than that of *from* (both to the men and the women, but *from one hundred to thousand), we remain agnostic regarding its category.
[9]Other constructions in the language do select the construction with *from*, e.g., *anywhere*: *anywhere *(from) 200 to 400 people*.
[10] The restriction in the occurrence of is achieved by making *x-to-y:whopping* a special subtype of *number* which is referred to only in the *whopping* construction. Other constructions which call for numbers (in particular, the noun-modification construction) will refer to a type which is incompatible with *x-to-y:whopping*.

8. References

Huddleston, Rodney and Geoffrey Pullum (2002). *Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Kim, Jong-Bok (2003). "Hybrid Agreement in English". In *Proceedings of the Fifth Annual Texas Linguistics Society*.

Michaelis, Laura A. (to appear). "Complementation by Construction". In M. Hauser et al., (eds.), *The Proceedings of the Thirty-second Annual Meeting of the Berkeley Linguistics Society: Parasession on Theoretical Approaches to Argument Structure*. Berkeley.

Sag, Ivan (2007). Sign-Based Construction Grammar: An informal synopsis. Manuscript, Stanford University.

Van Eynde, Frank (2006). "NP-Internal Agreement and the Structure of the Noun Phrase", *Journal of Linguistics* 42(1), 139–186.

Michael Ellsworth
infinity@icsi.berkeley.edu

Russell Lee-Goldman
rleegold@berkeley.edu

Russell Rhodes
russell_rhodes@berkeley.edu

Department of Linguistics
1203 Dwinelle Hall
University of California, Berkeley
Berkeley, CA 94720