

A Note on the Syntax of Numerical Bases

Richard S. Kayne
New York University

July 2005

1. Introduction (English)

The English number system has a base ten in the familiar sense that its multiplicative numerals are formed using powers of ten:

(1) There were four hundred/four thousand/four million people at the conference.

Ten itself does not have this possibility in a regular way:

(2) *There were four ten people at the conference.

What one has of course is:

(3) There were forty people...

in which at least the *-t-* (or perhaps *-ty*) is plausibly taken to be a reduced form of ten. That ten is a base and seven, for example, is not shown on the one hand by the impossibility of:

(4) *There were four seven people...

and on the other by the absence of any irregular counterpart to (4):

(5) *There were forsevy people...

Nor are there any multiplicative numerals based on powers of seven that would parallel *four hundred* or *four thousand*, etc.

In a similar way English additive numerals can involve powers of ten:

(6) two hundred and five

(7) three thousand five hundred and two

or the irregular forms based on ten itself just mentioned:

(8) forty-six

or ten alone, as in:

(9) sixteen

where *-teen* is straightforwardly taken to be a form of ten. But there are no additive numerals based, for example, on seven:

(10) *sixseven

Less often discussed as a reflection of the base status of ten is the following contrast:

(11) There were tens of thousands of people at the conference.

(12) *There were sevens of thousands of people...

In my English *tens of* is fully natural only with *thousands*, *millions*, *billions*, etc.:

(13) ?There were tens of mistakes in your paper.

But there is still a sharp contrast with:

(14) *There were sevens of mistakes in your paper.

and similarly for the other numerals below ten, all of which act like seven here.

The higher powers of ten are fully acceptable in sentences like (13):

(15) There were hundreds/thousands/millions of mistakes in your paper.

Other numerals higher than ten are not possible, for example:

(16) *There were twelves/fifteens of mistakes in your paper.

(17) *There were fifties of mistakes in your paper.¹

Although there is no doubt that ten is a numerical base in English, it may be that English in an extremely limited way allows twelve (in an irregular form *dozen* or perhaps *doz-*) to act as a base:

(18) There were dozens of mistakes in your paper.

If so, then English here would be acting a bit like French, which has ten as its normal base but in the case of eighty has:

(19) Elle a écrit quatre-vingt articles. ('she has written four-twenty articles')

in which twenty behaves as base for a multiplicative numeral.

In this paper, I will be interested in the question of how the language faculty treats the notion of numerical base. The proposals I will make will not require that a given language have exactly one base, nor will they care what number a given language has as its base. It may be that the choice of base and the choice of number of bases fall

outside the language faculty proper.² At the same time, I will argue that the language faculty has a very specific way of treating a numerical base (whatever its value and whether or not it is unique).

2. French

French is like English in allowing (15):

(20) Il y avait des centaines d'erreurs dans votre papier. ('it there had of-the hundreds of errors in your paper')

(21) ...des milliers d'erreurs... ('...of-the thousands...')

French also fully allows (13):

(22) Il y avait des dizaines d'erreurs dans votre papier. ('...of-the tens...')

At the same time, French differs from English in certain ways.

One is that *dizaines*, *centaines*, and *milliers* are preceded here by the so-called partitive article *des* (more precisely, by the preposition *de* ('of') followed by (a reduced form of) the definite article). In (13) and (15) no article at all precedes *tens*, *hundreds*, or *thousands*. This is part of a more general difference between the two languages concerning bare plurals, allowed by English but generally disallowed in French. This difference, though important in other ways, is not particularly relevant to what follows and I will not pursue it.

A second difference between English and French is that French has singular counterparts to (20)-(22):

(23) Il y avait une centaine d'erreurs dans votre papier. ('...a hundred of errors...')

(24) ...un millier d'erreurs... ('...a thousand of...')

(25) ...une dizaine d'erreurs... ('...a ten of...')

These are lacking in English:

(26) *There were a hundred of errors in your paper.

(27) *...a thousand of errors...

(28) *...a ten of errors...

Of course English has, without *of*:

(29) There were a hundred/a thousand errors in your paper.

But (29) conveys a precise numerical quantity, whereas (23)-(25) convey approximate numerical quantities much as (20)-(22), (15) and (13) do.

A third difference is that while English limits (13)/(15) to powers of ten (apart from *dozens*), French allows a wider range, in particular in the singular. Thus, in addition to (23)-(25), French has:

(30) Il y avait une quinzaine d'erreurs dans votre papier. ('...a fifteen+-aine of...')

(31) ...une vingtaine d'erreurs... ('...a twenty+-aine of...')

and similarly for *une trentaine d'* ('a 30+-aine of'), *une quarantaine d'* ('a 40+-aine of'), *une cinquante d'* ('a 50+-aine of') and *une soixantaine d'* ('a 60+-aine of').³

A fourth difference, the most important for what follows, is that all the French forms cited contain a visible suffix distinct from the numeral itself. For example *vingtaine* in (31) is transparently related to the numeral *vingt* (20), and similarly, with only minor irregularities, for:

(32) dix/dizaine (10)

(33) douze/douzaine (12)

(34) quinze/quinzaine (15)

(35) trente/trentaine (30)

(36) quarante/quarantaine (40)

(37) cinquante/cinquantaine (50)

(38) soixante/soixantaine (60)

(39) cent/centaine (100)

(40) mille/millier (1000)

With the exception of *millier*, all of these contain the suffix *-aine* preceded by the associated numeral. (The suffix in *millier* is *-ier*.) This suffix appears in the plural examples (20)-(22), as well as in the singular examples.

3. English and French

English *tens*, *hundreds* and *thousands*, while strongly parallel to (20)-(22) in interpretation and in being followed by a preposition (*of* in English, *de* in French), differ from the French forms in showing no visible suffix (apart from plural *-s*). This parallelism in interpretation (combined with the difference concerning *of* between *hundreds/thousands of books* and *a hundred/a thousand books*) led me in earlier work - cf. Kayne (2005, §3.1) - to propose that the English forms *tens*, *hundreds* and *thousands* actually do contain a suffix comparable to French *-aine*, though in English that suffix is unpronounced. Writing the unpronounced suffix as *-AINE*, we have, for example, for *tens of thousands of books* and *hundreds of books*:

(41) ten -AINE -s of thousand -AINE -s of books

(42) hundred -AINE -s of books

Multiplicative numerals like *two hundred* or *three thousand* differ from the preceding in not being intrinsically approximative and in not allowing *of*:⁴

(43) two hundred (*of) books

(44) three thousand (*of) people

The French counterparts of these do not show a preposition, either:

(45) deux cents (*de) livres

(46) trois mille (*de) personnes

and in that respect contrast with the approximative examples of (20)-(25) and (30)-(31), all of which do contain *de*.

The difference in interpretation (approximative vs. not) and the difference concerning *of/de* led me in the earlier work mentioned to take the position that *hundred* and *thousand* (and *cents* and *mille*) in (43)-(46) were not followed by any suffix (apart from the plural -s in French *cents*), contrary to the analysis of (what I will informally call) approximatives illustrated in (41) and (42).

4. Similarities between approximatives and multiplicative numerals

That decision overlooked some similarities, however. One is seen starting with the pair:⁵

(47) There were nine/eleven hundred linguists at the conference.

(48) *?There were ten hundred linguists at the conference.

This contrast resembles:

(49) There were tens of thousands of people at the demonstration.

(50) ??There were tens of hundreds of people at the demonstration.

As a multiplicative numeral, *?*ten hundred* is quite unnatural. To my ear that unnaturalness is shared to a significant degree by ??*tens of hundreds*. The same holds even more sharply for me in:⁶

(51) They have nine hundred thousand dollars in their bank account.

(52) *They have ten hundred thousand dollars in their bank account.

paralleled by the following contrast:

(53) They must have tens of thousands of dollars in their bank account.

(54) *They must have tens of hundreds of dollars in their bank account.

A second similarity between multiplicative numerals and the approximatives under discussion holds with respect to the requirement that in a multiplicative numeral a lower numeral must precede a higher one (in both English and French):

(55) They bought three hundred books.

(56) They have five hundred thousand dollars.

These are the only well-formed orders:⁷

(57) *They bought hundred three books.

(58) *They have five thousand hundred dollars.

It turns out that a similar property holds for approximative expressions:

(59) They must have tens of thousands/hundreds of thousands of dollars in their account.

Here the lower precedes the higher. The reverse is not possible:

(60) *They must have thousands of tens of dollars in their account.

(61) *They must have thousands of hundreds of dollars in their account.

The question is how best to express the similarity between (60)/(61) and (57)/(58). (Part of the answer must rest on the fact that (59), while approximative, is also multiplicative.)

The same question arises for a third similarity between multiplicative numerals and (multiplicative) approximatives:

(62) *They have five hundred hundred dollars in their account.

(63) *They have five thousand thousand dollars in their account.

The deviance of these cases of repetition of the same numeral seems to extend to:

(64) *They must have hundreds of hundreds of dollars in their account.

(65) *They must have thousands of thousands of dollars in their account.

Again, the fact that (59) is multiplicative in interpretation seems crucial, since the following, with *and* instead of *of* and an additive interpretation, are fine:

(66) They must have hundreds and hundreds of dollars in their account.

(67) They must have thousands and thousands of dollars in their account.

Also acceptable with an additive interpretation (and with *upon*) are:

(68) They must have hundreds upon hundreds of dollars in their account.

(69) They must have thousands upon thousands of dollars in their account.

A fourth set of facts that seems to suggest a closer relation between multiplicative numerals and multiplicative approximatives than one might have thought involves *several*, which in English can combine with *hundred* and *thousand* in a way resembling the way in which multiplicative numerals are formed:⁸

(70) They have several hundred dollars in their account.

(71) They have several thousand dollars in their account.

French *plusieurs* seems in general like an excellent match for English *several*. Yet it cannot combine with *cent* or *mille* at all:

(72) *Ils ont plusieurs cent dollars dans leur compte.

(73) *Ils ont plusieurs mille dollars dans leur compte.

Close to (70) and (71), instead, are:

(74) Ils ont plusieurs centaines de dollars dans leur compte.

(75) Ils ont plusieurs milliers de dollars dans leur compte.

whose closest word-for-word counterparts in English are less natural than (70) and (71):

(76) ?They have several hundreds of dollars in their account.

(77) ?They have several thousands of dollars in their account.

The fact that English (70) and (71), with *hundred* and *thousand* in the manner of multiplicative numerals, transpose into French yielding (74) and (75), which contain the suffix-bearing (and plural) *centaines* and *milliers*, suggests that (70) and (71) contain nominal suffixes, too:

(78) ...several hundred NSFX dollars...

(79) ...several thousand NSFX dollars...

where NSFX is an informal abbreviation for nominal suffix. This proposal turns out to receive convergent support from Romanian.

5. Romanian.

As a (further) way of reaching an analysis that expresses what multiplicative numerals and multiplicative approximatives (the ones containing more than one numeral, as in *hundreds of thousands of books*) have in common, let us turn to Romanian,⁹ which displays a phenomenon that is intriguingly, though not exactly, similar to what is found in the English contrasts seen earlier, such as in:

(80) a hundred (*of) books

(81) hundreds *(of) books

Also:¹⁰

(82) ten (*of) books

(83) tens *(of) books

What is notable about Romanian is that we find something comparable (concerning the presence or absence of a preposition) entirely within the domain of ordinary numerals, in particular with ten and its simple multiples (20, 30, etc.).

Ten in Romanian is followed directly by the (plural) noun, with no preposition:

(84) zece caiete ('ten notebooks')

and in that sense resembles (82). Thirty in Romanian is more transparent than in English (or French) and has the form 'three tens'. A following (plural) noun must be separated from Romanian 30 by the preposition *de*:

(85) treizeci de caiete ('three-tens of notebooks' = '30 notebooks')

and in that sense 30 in Romanian acts like (83) in English, despite the fact that (83) is not an example of an ordinary numeral.

The generalization might seem to be that the presence of the preposition (*of* or *de*) is keyed directly to the plural form of *tens* and *zeci*. But why would that be? Let me instead propose that what is central, or at least more central, is the unpronounced nominal suffix mentioned earlier as being present in (83) - cf. the discussion of (20)-(42) - as illustrated in:

(86) ten -AINE -s of books

with the absence of *of* in (82) then correlating with the absence there of any such suffix.

Generalizing to Romanian, we can attribute the contrast concerning *de* between (84) and (85) to the presence of a nominal suffix in (85) vs. the absence of any such suffix in (84). We should consequently think of (85) as:

(87) trei zec- NSFX -i de caiete

where NSFX is the nominal suffix in question (and *-i* the plural morpheme). If we now ask why (85), but not (84), contains this suffix, the generalization would seem to be:¹¹

(88) In Romanian, *zec-* is followed by NSFX whenever *zec-* is itself multiplied by a preceding numeral.

6. UG and numerical bases

What (88) says is that in a multiplicative numeral based on 10 in Romanian, 10 must be associated with a nominal suffix. The facts of Romanian are consistent with generalizing this to all multiplicative numerals in Romanian. But let me jump immediately to a much broader proposal:

(89) In all languages, the multiplicand of a multiplicative numeral must be associated with a nominal suffix.

In a numeral like three hundred, 'hundred' is the multiplicand, so (89) requires:

(90) three hundred NSFX...

In a numeral like three hundred thousand, there are two multiplicands, 'hundred' and 'thousand', so (89) plausibly leads to:¹²

(91) three hundred NSFX thousand NSFX...

Note that (89) gives us a way of expressing syntactically the impossibility of:

(92) *three seven books

if we grant that only a base numeral (or one of its powers) can combine with NSFX. Put another way, the choice of a base numeral for a given language amounts to that language allowing that numeral to combine with NSFX. Since seven in English is not a base numeral, it cannot combine with NSFX, so (92) violates (89).¹³

7. Prepositions

A generalization concerning prepositions can now be stated as follows:

(93) If in a given language there are two forms involving numerals or approximatives that differ in that one contains a nominal suffix and the other not, then either

- a) both have a preposition or
- b) neither has a preposition or
- c) only the one with the nominal suffix has a preposition

Case (c) corresponds to Romanian numerals as in (84) vs. (85); the latter has a nominal suffix as indicated in (87), the former does not. Case (b) corresponds to English *three books* vs. *three hundred books* - although only the latter has NSFX as imposed by (89),¹⁴ neither contains *of*.

The English contrast between *three hundred books* and *hundreds of books*, while not falling under (93), indicates that the NSFX associated with numerals calls for the presence of a preposition less strongly than the *-AINE* suffix associated with approximatives.¹⁵ This, along with the perhaps related fact (if it is one) that NSFX is pronounced less widely than *-AINE*, remains to be better understood. What (93) excludes is the possibility of a language just like Romanian but with *de* appearing with 10 and not with 30.¹⁶

8. More on NSFX

The question arises as to whether NSFX is ever pronounced. Within the Romance and Germanic families, I can think of two candidates. One is Italian *-ant-* as in:

(94) cinquanta; settanta; ottanta; novanta

which correspond to 50, 70, 80 and 90 and are readily decomposable as:

(95) cinqu+ant+a etc.

where the first part is the numeral 5, 7, 8, or 9 (less its final vowel), the second part *-ant-* and the third part the most usual Italian ending for a feminine singular noun. (30, 40 and 60 in Italian are almost as transparent as these; 20 is not.) It is possible, of course, that *-ant-* here should be analyzed as a suppletive variant of ten (*dieci*, in Italian).

Alternatively, the proper analysis might be:

(96) cinqu + TEN + ant + a

with TEN unpronounced and *-ant-* the spelling out of NSFX. (The same might hold for French, though the forms are less regular than in Italian.)

The second candidate is found in English *million*, *billion*, *trillion*. Given this series (along with the less usual *quadrillion*, *quintillion*), it is tempting to factor out *-illion* as a possible realization of NSFX. Alternatively, *-illion* might have some precise numerical content, though that might be hard to square with its appearance in *zillion*, which lacks a precise value.

These larger numerals raise other questions, e.g. in French:

(97) trois cents (*de) livres ('300 books')

- (98) trois mille (*de) livres ('3000 books')
- (99) trois millions *(de) livres ('3 million books')
- (100) trois milliards *(de) livres ('3 billion books')

Unlike the French counterparts of *hundred* and *thousand*, the French *million* and *milliard* require *de*. From the present perspective, this suggests the presence in the case of *million* and *milliard* of a suffix not present with *cent* or *mille*. It might be that (99), e.g., is:

- (101) trois million NSFX -AINE₂ -s de livres

with an unpronounced -AINE₂ in addition to NSFX, where -AINE₂ is a weaker version of -AINE. This would amount to claiming that the larger single-word numerals like *million* are intrinsically associated with approximation.

The presence of -AINE₂ with *million* (and *milliard*) in French would then correlate with the contrast:

- (102) cent livres
- (103) mille livres
- (104) un million de livres

In addition to differing from *cent* and *mille* with respect to *de*, French *million* also differs from them in that *million* requires *un* ('a/one') preceding it when there is no other preceding numeral.

A further difference in French between *cent* and *mille* on the one hand and *million* on the other is that the full-fledged approximatives of the first two show a suffix lacking with the approximative corresponding to *million*:

- (105) des centaines de livres ('of-the hundreds of books')
- (106) des milliers de livres ('of-the thousands of books')
- (107) des millions de livres ('of-the millions of books')

Centaine contains *-aine* and *millier* contains *-ier*, but there is no corresponding suffix in (107). This might be related to the very fact that *million* as an ordinary numeral (cf. (104)) is already associated with -AINE₂.

In Romanian, the cutoff that corresponds to (102)-(104) is between 10 and 100:

- (108) zece caiete ('ten notebooks')
- (109) o suta de caiete ('a hundred of notebooks')

It may be that Romanian associates -AINE₂ with the single-word powers of ten beginning with 100, while French 'waits until' 1,000,000.

9. Constituent structure.

The postulation of NSFX via (89) is, strictly speaking, independent of the question of the constituent structure of multiplicative numerals or multiplicative approximatives. Yet the question needs to be addressed. Ionin and Matushansky (2005) propose:¹⁷

- (110) three [hundred books]

whereas (89) seems to fit more naturally with:

- (111) [three hundred] books

in which there is a constituent that corresponds to 'multiplicative numeral'.

There is some evidence suggesting that (111) is closer to being correct than (110). Given the arguments presented earlier (see (47)-(77)) in favor of a substantial parallelism between multiplicative numerals and multiplicative approximatives (which we now see to rest at least in part on the presence in each of a nominal suffix, either NSFX or -AINE),¹⁸ evidence from the latter is relevant to the analysis of the former.

Consider then:

- (112) They have hundreds of thousands of linguistics books.

Preposition-stranding under topicalization gives a sharp contrast:

- (113) (?)Linguistics books they have hundreds of thousands of.
- (114) *Thousands of linguistics books they have hundreds of.

The strong deviance of (114) suggests that *thousands of linguistics books* in (112) is not a constituent, and that we rather have:¹⁹

- (115) [hundreds of thousands] of linguistics books

With ordinary numerals, topicalization in English is not natural (for me), though there is still a sharp contrast seen in:

- (116) ??Linguistics books they have three hundred.
- (117) *Hundred linguistics books they have three.

supporting (111) over (110). Again starting from:

- (118) They have three hundred linguistics books.

we find a similar contrast involving:

(119) They have three hundred linguistics books and we have three hundred, too.
as opposed to:

(120) *They have three hundred linguistics books and we have three, too.
(This similarity in contrasts suggests that (119) involves movement of unpronounced ‘linguistics books’.)

10. Why is NSFX needed?

Assume that (111) is on the right track. Making NSFX explicit in it, we have:

(121) [three [hundred NSFX]] books

where ‘[three [hundred NSFX]]’ is in the Spec of some functional head (cf. note 17). Comparing this with:

(122) three books

where ‘three’ itself is in a corresponding Spec position, it is natural to conclude that in (121) ‘three’ itself is in a Spec position within ‘[three [hundred NSFX]]’.

Since in (122) ‘three’ is paired with a NP (‘books’) and since by hypothesis NSFX is nominal, we see that in (121), in parallel fashion, ‘three’ is also paired with a NP, namely ‘hundred NSFX’. This gives us a way of excluding:

(123) *[three hundred] books

i.e. of enforcing the presence of NSFX, namely by taking numerals to require being paired with a NP²⁰ and by taking ‘hundred’ not to be one.²¹

As for the interpretation of NSFX, pairs like the following come to mind:

(124) They entered the room in threes.

(125) They entered the room in groups of three.

as do, especially:

(126) Counting these grains of rice would go faster if you counted them by threes/groups of three/sets of three.
where the grouping comes into being solely as a function of the counting. In other words, NSFX may be equivalent to (unpronounced) SET.

11. Conclusion.

Multiplicative numerals like *three hundred* have more in common with multiplicative approximatives (like *hundreds of thousands*) than one might have thought. Each depends on the presence of a nominal suffix. In the case of multiplicative numerals that (unpronounced) suffix is associated with the numeral base (or its power) and is probably interpreted in a way akin to *set*. A numeral base (or its power) can be multiplied by another numeral only if combined with such a nominal suffix.

References:

Barbiers, S. (2004)%

Corver, N. (2001) “On predicate numerals,” in T. van der Wouden and H. Broekhuis (eds.) *Linguistics in the Netherlands 2001*, John Benjamins, Amsterdam, 65-67.

Ionin, T. and O. Matushansky (2005) “1001 Nights: The Syntax and Semantics of Complex Numerals,” ms., USC and CNRS.

Kayne, R.S. (2003a) “Silent Years, Silent Hours”, in L.-O. Delsing et al. (eds.) *Grammar in Focus. Festschrift for Christer Platzack. Volume 2*, Wallin and Dalholm, Lund, 209-226 (reprinted in Kayne (2005b)).

Kayne, R.S. (2003b) “Antisymmetry and Japanese”, *English Linguistics*, 20, 1-40 (reprinted in Kayne (2005b)).

Kayne, R.S. (2005a) “Some Notes on Comparative Syntax, with Special Reference to English and French” in G. Cinque and R. Kayne (eds.) *Handbook of Comparative Syntax*, Oxford University Press, New York (reprinted in Kayne (2005b)).

Kayne, R.S. (2005b) *Movement and Silence*, Oxford University Press, New York.

Zweig, E. (2005) “Nouns and Adjectives in Numeral NPs,” *NELS Proceedings*.

¹ This example contrasts with the possible:

- (i) John is in his fifties.
- (ii) John was born in the fifties.

both of which arguably contain (unlike (17)) an unpronounced YEAR(S) - cf. Kayne (2003a).

Apparently closer to (17) is:

- (iii) Her articles must number (somewhere) in the fifties.

though here (again unlike (17)) there may be an unpronounced pronominal N whose antecedent is *articles*:

- (vi) ...(somewhere) in the fifty N -s

² Cf. Ionin and Matushansky (2005, §9.2.1).

³ A plural example provided by Jean-Yves Pollock is:

i) La dernière fois...Jean a acheté des dizaines, que dis-je?, des vingtaines, peut-être des trentaines de CD! ('the last time...John has bought of-the tens, what say I, of-the twenties, may-be of-the thirties of CD')

⁴ Apart from cases like *two hundred of these books*, which are not relevant to the present discussion.

⁵ The reason for these contrasts is not entirely clear. The deviance of **ten hundred* and *??tens of hundreds* could plausibly be related to the possibility of replacing them with *thousand* and *thousands*. Yet *eleven hundred* in (47) is well-formed despite being replaceable by (the also well-formed) *one thousand, one hundred*.

⁶ Left open is why (i) is less good than (47):

- i) ??They have eleven hundred thousand dollars in their bank account.

⁷ Irrelevantly to the present discussion, (57) and (58) are possible (in French) as examples of additive numerals:

- i) Ils ont acheté cent trois livres. ('103')
- ii) Ils ont cinq mille cent dollars. ('5100')

⁸ The parallel is supported by an observation in Ionin and Matushansky (2005), who note:

- i) several hundred/*forty books

just like:

- ii) three hundred/*forty books

⁹ I am grateful to Oana Savescu-Ciucivara for all the Romanian data.

¹⁰ Strictly speaking, (83) is not perfect for me (cf. (13)), though (11) is. In any case, the contrast concerning *of* between (82) and (83) is very sharp.

¹¹ This formulation assumes either that *zec-* in (84) is not multiplied by 1, or else that 1 is not a numeral - on the latter possibility, cf. Barbiers (2004).

¹² This conclusion concerning (91) is connected to the question of constituent structure to which I return briefly below. An alternative that seems less plausible is:

- i) three [hundred thousand NSFX] NSFX

¹³ Similarly for:

- i) *three seven thousand books

¹⁴ I am assuming that numerals are associated with NSFX only when required to be by (89).

¹⁵ The Romanian contrast between (84) and (85) means that if low numerals are accompanied by unpronounced NUMBER (cf. Zweig (2005)), then NUMBER is to NSFX as NSFX is to -AINE, with respect to how strongly the presence of a preposition is required.

As Zweig notes, the question of the categorial status of the simple numeral itself (*three* or *seven*, for example) is a separate one.

¹⁶ Another factor bearing on the presence of a preposition is whether or not the NP is preceded by an overt determiner. Elabbas Benmamoun (p.c.) tells me that in Moroccan Arabic the numeral two allows (the counterpart of) either ‘two books’ or ‘two of the books’ (both with an indefinite interpretation), with ‘of’ in the latter apparently induced by ‘the’.

The presence of a definite article in Moroccan Arabic in what corresponds to an indefinite is almost certainly closely related to French:

i) Ils ont acheté des livres. (‘they have bought of-the books’)

with a definite article yet with an interpretation akin to *They bought (some) books*.

¹⁷ In addition to the facts discussed below, their proposal would seem to have a problem with Romanian numerals followed by *de*. As Corver (2001) has noted, the presence of *de* fits naturally into an analysis in which the (complex) numeral moves to the Spec of *de*.

¹⁸ An apparent discrepancy is:

i) hundreds of thousands of books

ii) *three hundred thousands of books

which might be related to:

iii) three hundred thousand books

iv) *hundreds of thousand books

in terms of syntactic ‘agreement in approximateness’ (cf. Ionin and Matushansky (2005)’s similar pragmatic principle, in an additive context), though that would have to distinguish (iv) and (v):

v) *a lot of thousand books

from:

vi) several/?many thousand books

A advantage of ‘agreement in approximateness’ would be (v) vs.:

vii) a couple of thousand books

On the additive side, such agreement does not seem necessary in:

viii) five-hundred some-odd books

¹⁹ This constituent structure would also appear to be supported by the contrast (for me):

i) ?tens of books

ii) tens of thousands of books

Similarly:

iii) Your books/published articles must number in the tens of thousands/?tens by now.

presumably with an unpronounced ‘books’.

²⁰ Or with some projection containing NP, in cases such as those mentioned in note 16.

²¹ Alternatively, by taking ‘hundred’ not to be of the right sort of NP. Relevant here is Ionin and Matushansky’s (2005) discussion of Case in Russian.

Whether the text idea could be made compatible with (110) is unclear.

Pierre Pica (p.c.) has suggested a link between the absence of higher numerals and the absence of sentential complementation, in some languages. If so, then the appearance of NSFX has something important in common with the nominalization of sentential complements proposed in Kayne (2003b, §4.6).

