Nouns and Adjectives in Numeral NPs*

Eytan Zweig

July 6, 2005

1 Introduction

It is generally taken to be a linguistic fact that every language features cardinal numerals as part of its syntactic system. It is not unreasonable, then, to assume that numerals are one place where cross-linguistic comparisons are especially fruitful and syntactic universals may be uncovered. However, once these systems are examined more closely, the picture becomes a lot more complicated, for it appears that numerals do not belong to a uniform syntactic category. While some languages feature numerals which show properties of adjectives, others feature numerals which appear nominal. Indeed, many languages show internal variation in the properties of numerals. However, this variation does not appear totally arbitrary: if a language has varying numerals, invariably lower numerals have adjectival properties and higher numerals appear more nominal (see Jespersen (1969), Corbett (1978), Hurford (1987)).

This paper examines evidence from the numeral systems of several unrelated languages, and argues that there is, despite the apparent variation, a single universal syntax for numeral phrases. Following the analysis of numerals proposed by Ionin and Matushansky (2004a,b) in conjunction with the proposals made in Kayne (2003, 2005), it claims that even when the numerals themselves are adjectival, the numeral phrase itself is nominal, with an unpronounced head nominal.

1.1 Terminology

Before proceeding in the paper, it is important to clarify some of the terminology involved. Throughout this paper, the term **numeral** will be used to refer to a sequence that describes a number. Examples of numerals are *two*, *seventy-three*, *five hundred*, and *five million, three-hundred thousand and twenty-nine*. Each morphological unit within such a sequence is a **numeral word**. *Two* and

*I'd like to thank Richard Kayne, Anna Szabolcsi, Rahul Balusu, Yael Fuerst, Tom Leu, Lisa Levinson, Laura Rimell, Kevin Roon and Oana Savescu-Ciucivara for their useful comments, suggestions and data for early versions of this paper. I'd also like to thank the participants at NELS 35, especially Ora Mathushansky, and at ConSOLE 13 for their helpful feedback.

million are numeral words. A numeral that consists of a single numeral word is called a **simple numeral**.

Numerals can be joined with other numerals in various ways to form **complex numerals**. The two most important and prevalent forms of complex numerals are **additive numerals**, wherein the values of two numerals are added to produce the value of the whole (e.g. *fifty-five*), and **multiplicative numerals**, wherein the values are multiplied (e.g. *two hundred*)[1]. Note that numeral formation is recursive and complex numerals can be used to assemble other numerals; for instance, the numeral *two hundred and five thousand* is a multiplicative numeral formed from multiplying *two hundred and five* - an additive numeral, which itself contains a multiplicative numeral - and the simple numeral *thousand*.

Finally, a note on notation. Throughout this paper, when the name of a numeral is spelled out, as in *thousand*, it refers to the actual string in the language in which it appears. The meaning, or value, of a numeral will always be given in numbers. So the numerals *three* in English and *shalosh* in Hebrew both mean 3.

2 Numerals as Nominal Modifiers

Ionin and Matushansky (2004a,b) argue that numeral words are nouns which act as nominal modifiers, taking NP complements.

To justify this claim, Ionin and Matushansky must first show that standard analyses of numeral structure are incorrect. First, they argue that numerals are not heads in either a determiner or dedicated NumP position. The most obvious argument is the very existence of complex numerals of arbitrary complexity. While it may be possible to argue that *twenty-two* is a frozen lexical head, this is very implausible for numerals such as *a million three hundred and five thousand, seventy six*.

Furthermore, in many languages, it is possible to see syntactic interactions within numerals. For instance, in Bantu languages, there is internal agreement within numerals. In Luganda[2], the number 20 is expressed by a phrase equivalent to "two tens". As seen in (1), *biri* ('two'), shows class agreement with *ma-kumi* ('ten'), and not *mi-dumu* ('jug'):

(1) emi-dumu ama-kumi a-biri
MI-jug MA-ten AGR_{MA}-two
'twenty jugs'

[1]Some languages, such as Yaruba, feature subtractive numerals. While I will not deal with them here, I assume that their structure resembles that of additive numerals and that the differences between them are mainly semantic. Even rarer are divisive systems, which I take to have a basic structure similar to multiplicative numerals. See Greenberg (1978) and Hurford (1975) for discussion of subtractive and divisive systems.

[2]See Ashton et al. (1954)

This would be difficult to explain if *ama-kumi a-biri* was a head, but predicted if *kumi* was a noun modifying *dumu* and in turn modified by *biri*. Note that nouns modfying other nouns in Luganda do not show agreement.

A second possibility that Ionin and Matushansky argue against is that numerals are phrasal, and are located in the specifier of NumP or QP. They bring as counter-evidence the fact that in languages such as Russian and Inari Sami, numerals assign case to the noun they modify. For instance[3]:

(2) Russian:

a. četyre šagA
four step-Pauc
'four steps'

b. šest' šagov
six step-Gen.pl
'six steps'

(3) Inari Sami:

a. kyehti päärni
two two child-Acc.sg
'two children'

b. čiččâm pärnid
seven child-Part.sg
'seven children'

In both these examples, the case of the head noun depends on the numeral involved. This makes it implausible, under the standard assumption that case is assigned by heads, that numerals are specifiers.[4]

Having shown that numerals can be neither heads nor phrases in specifier positions, Ionin and Matushansky (2004a,b) go on to argue that (complex) numerals are not constituents, but instead are built incrementally. Multiplicative numerals, such as the one in *two hundred dollars* have the structure shown in (4):

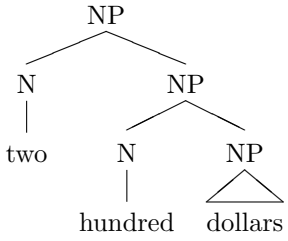
[3]Examples and glosses from Ionin and Matushansky (2004b)

[4]Ionin and Matushansky (2004a,b) provide many arguments beyond the ones listed here, as well as a compositional semantics for their numeral system, and an account of modified numeral constructions:

i a healthy twelve performances

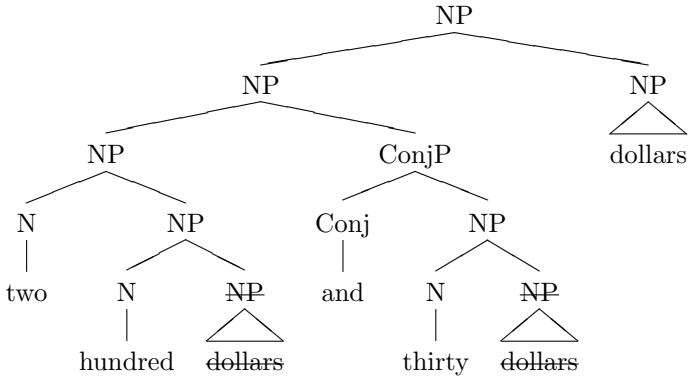
Readers are directed to their papers for the full story.

(4) two hundred dollars



Additive numerals, on the other hand, are formed from conjunctions (either overt or covert) of two numerals, with the head noun then undergoing right-node raising, as in (5)[5]:

(5) two hundred and thirty dollars



Evidence for right node raising can be found in several languages, including Biblical Hebrew and some Bantu languages, where the raising is optional, allowing reduplication of the noun phrase. Biblical Hebrew is especially useful for illustrating this, as it allows the noun to occur in every member of the conjunction:

(6) meah shana ve-esrim shana ve-sheva shanim
hundred year and-twenty year and-seven years
'one hundred and twenty-seven years' (Genesis 23:1)

A similar phenomenon can be seen in Luvale[6]:

(7) mikoko makumi atanu na-mikoko vatanu
sheep ten five and-sheep five
'fifty-five sheep'

[5]This tree is an informal representation for expository purposes, and should not be taken to represent a true theory of the structures involved in right node raising.

[6]See Horton (1949)

3 Adjectival and Nominal Numerals

The analysis given above treats all numerals as nouns. As mentioned in the introduction, however, the actual situation seems more complicated. In many languages, the simple numerals do not show the properties of a single syntactic category. Instead, lower numerals tend to share syntactic and morphological properties with adjectives, while higher numerals tend to exhibit properties associated with nouns (see Corbett (1978)).

Several examples follow. In many Bantu languages, numerals lower than a certain threshold (often 5 or 10, though much variation exists) agree with the noun they modify, featuring adjectival or enumerative agreement prefixes, as in example (8) from Luganda. Higher numerals do not agree, instead featuring their own nominal class prefixes (9):

(8) emi-dumu e-biri
MI-jug AGR_{MI}-two
'two jugs'

(9) emi-dumu mu-sanvu
MI-jug MU-seven
'seven jugs'

Example (1) (repeated as (10)) shows that a higher numeral triggers agreement on an adjectival numeral modifying it, further evidence that it shows nominal behavior:

(10) emi-dumu ama-kumi a-biri
MI-jug MA-ten AGR_{MA}-two
'twenty jugs'

Further examples for this pattern can be seen in English, where, starting with *hundred*, numerals can appear in plural form in partitive constructions (11), can take determiners (12) and can be modified by other numerals (13):

(11) a. hundreds of boys
b. * threes of boys

(12) a. a/several hundred boys
b. * a/several three boys

(13) a. four hundred boys
b. * four three boys

In Modern Hebrew, numerals up to 19 agree in gender with the head noun (14b), but higher numerals do not (14a):

(14) a. shloshim yeladim/yeladot
thirty boys/girls
'thirty boys/girls'

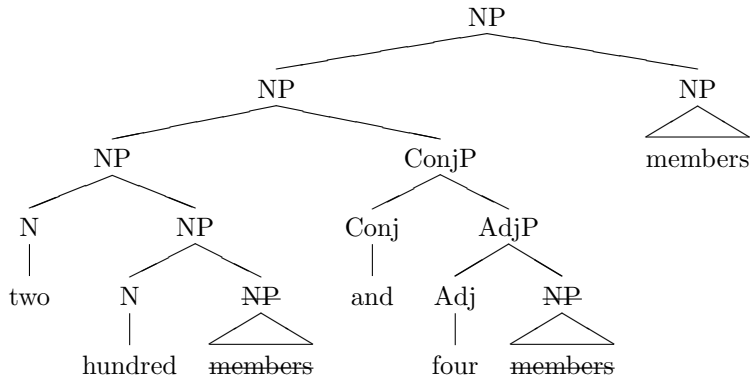
b. shlosha yeladim/*yeladot
three-masc boys/*girls
'three boys'

Thus, an account that classifies all numerals as nouns fails to account for much of this data. One obvious hypothesis is that this does not matter; the syntax is exactly as given by Ionin and Matushansky (2004a,b), the only difference being in the categorical specification of the numeral. However, this analysis does not seem supported by the facts. For instance, in English, no one would question the grammaticality of (15):

(15) two hundred and four members

According to the right-node raising analysis of additive numerals, this DP must have the following structure:

(16)



However, it is not generally possible in English to have right node raising from a conjunction of an adjectivally modified NP and a nominally modified NP:

(17) a. Our neighborhood has big and small houses
b. Our neighborhood has brick and wood houses

But not:

(18) a. * Our neighborhood has big and wood houses
b. * Our neighborhood has brick and small houses

A second indication that this proposal is not sufficient can be found in Hebrew (both Biblical and Modern), where, though low numerals exhibit some adjectival properties as shown above, all numerals greater than one precede the head noun, in a structure reminiscent of the construct state[7], a position restricted to NPs, while adjectives follow the head noun:

[7]See Zadka (2001) for discussion of whether this is an actual construct state.

(19) (*adumim) batim adumim
(*red) houses red
'red houses'

(20) shnei batim (*shnaim/*shnei)
two houses (*two)
'two houses'

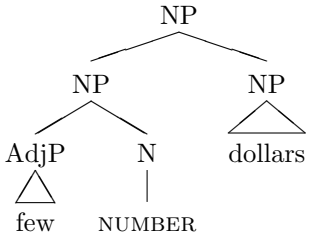
(21) ba'al batim (*ba'al)
owner houses (*owner)
'owner of houses'

It seems, then, that there is a paradox here. On the one hand, low numerals appear to be adjectives. On the other hand, they seem to have the distribution of NPs. A solution to this apparent contradiction can be found in the recent analyses of quantity words such as *few* and *many* proposed by Kayne (2005).

4 Numerals and number

Kayne (2005) provides a detailed argument that shows that *few* and *many* are adjectives, but instead of modifying nouns directly, they modify an unpronounced noun which he names NUMBER, and in turn, the whole NUMBER NP modifies the noun, as exemplified in (22):

(22) few dollars



This structure explains, among other things, why, even though determiners such as *every* and *a* normally modify only singular NPs, they can appear in plural NPs if they are modified by *few*, but not other adjectives:

(23) a/every few books

(24) * a/every good books

Adjectival numerals, too, are compatible with *every*:

(25) every three sentences

Nominal numerals are compatible with both every and a:

(26) a/every hundred dollars

This data makes extending Kayne's proposal to numerals a natural step. In this view, all numerals are NPs. The low, adjectival numerals such as *three* take the form in (27a), while higher numerals feature an overt noun, as in (27b).

(27) a. $[_{\text{NP}}\ [_{\text{AP}}$ three] NUMBER]
b. $[_{\text{NP}}\ [_{\text{AP}}$ three] thousand]

This explains the examples above. In (26), the availability of every is licensed by the singular nominal numeral *hundred*; yet, given that *three* is an adjective, no such overt nominal appears in (25). Postulating a covert noun provides an explanation for the sentence's acceptability, just like it does in (23).

This leaves open the question of why (28) is bad:

(28) * a three sentences

However, note the following pattern with an overt nominal numeral:

(29) (*a) two hundred books

(30) (a) few hundred books

Here, too, *a* is incompatible with a numeral but permissible with *few*. This may be because *a* is really a variant of *one*, and therefore incompatible with other numerals for semantic reasons.

4.1 Evidence from Hebrew

In Hebrew, there is even stronger evidence for the above among a certain class of nouns. While most nouns require the plural form with any numeral, these nouns can occur in the singular with a nominal numeral, but not with an adjectival numeral[8]:

(31) shlosh meot yom/yamim
three hundreds day/days
'three hundred days'

[8]Zamparelli (2004) provides an alternative explanation of *every X* constructions that would render the argument below invalid. He, too, postulates a covert noun; however, this noun is not part of the numeral, but rather it represents a linear sequence measured by the lexical noun:

i every two days TIME

this argument relies on the lexical noun in these constructions being a unit of measurement. The Hebrew data are a problem for this, as the class of nouns that have this property is only loosely defined, and varies among different speakers. While most of its members are descriptions of units of time, measurement, and currencies, it is not restricted to these (among many speakers *ish* ('person') belongs to this class), nor does it contain all of them (no speaker accepts *shnia* ('second') as a member). For a brief discussion on possible characteristics of this class, see Zadka (2001).

(32) shalosh *yom/yamim
three *day/days
'three days'

When further modified by *kol* ('every'), however, an interesting fact emerges. If the nominal numeral is singular, the head noun can be either singular or plural (33). If it is plural, as is the case when forming the multiples of 100 from 300 onwards[9], the lexical noun itself must be singular (34):

(33) kol me'a yom/yamim
every hundred day/days
'every hundred days'

(34) kol shlosh meot yom/*yamim
every three hundreds day/*days
'every three hundred days'

However, with adjectival numerals, *kol* can (and must) co-occur with a plural lexical noun:

(35) kol shalosh *yom/yamim
every three *day/days
'every three days'

If, as (33) and (34) indicate, there must be a singular noun somewhere in a phrase to license kol, (35) must contain one as well. This, of course, would be NUMBER:

(36) kol shalosh NUMBER *yom/yamim
every three NUMBER *day/days
'every three days'

4.2 Evidence from Luvale

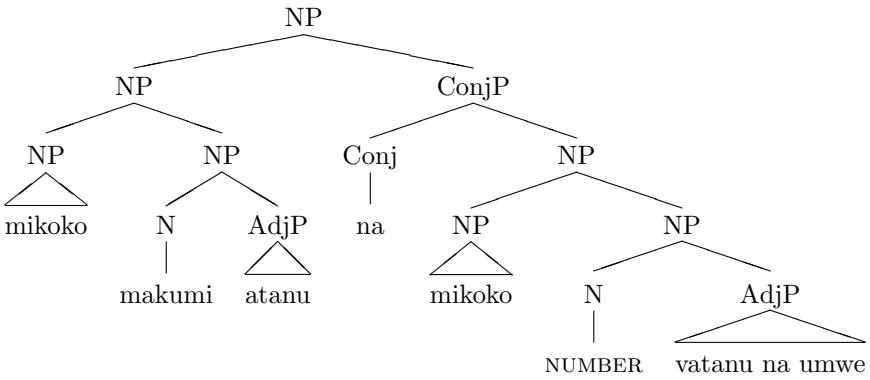
Luvale contains a different kind of evidence for the NUMBER proposal. As mentioned above, in Luvale additive numerals, the lexical noun is reduplicated between different numeral conjuncts, as right-node raising does not take place. However, the situation is more complex than in Biblical Hebrew, as there are two kinds of additive numerals. While Luvale is a base 10 language, meaning that multiplicative numerals are multiples of 10 and that these multiples are joined into additive numerals, numbers between 6-9 are also additive, formed by a conjunction of 5 and the appropriate numeral. However, the head noun does not appear within this conjunction[10]:

[9] 200 is represented by a dual form rather than by a multiplicative numeral
[10] See Horton (1949)

(37) mikoko makumi atanu na-mikoko vatanu na-umwe
sheep ten five and-sheep five and-one
fifty-six sheep'

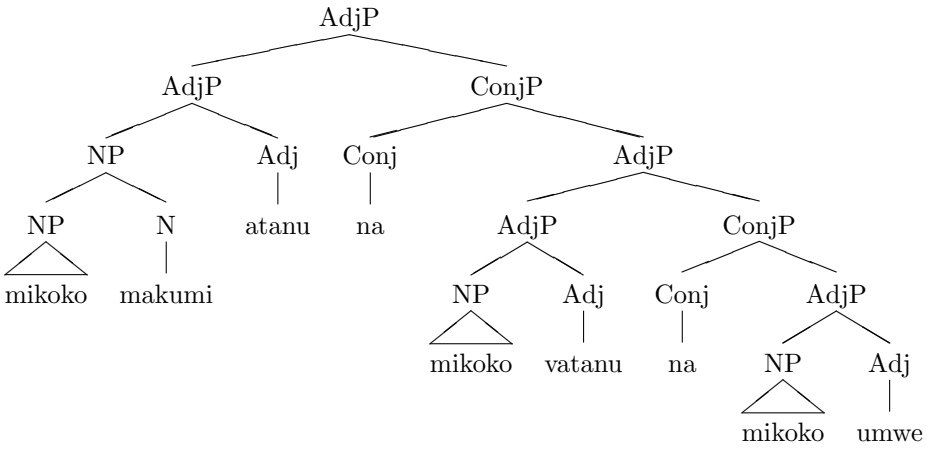
This data confirms to the predictions of the NUMBER hypothesis, which posits the structure in (38) for the DP in question:[11]:

(38) mikoko makumi atanu na-mikoko vatanu na-umwe



On the other hand, Ionin and Matushansky (2004a,b), which posits identical structures for adjectival and nominal numerals, would require a three-way conjunction, as in (39), which would incorrectly allow *mikoko* to appear before *umwe*:

(39) * mikoko makumi atanu na-mikoko vatanu na-mikoko umwe



[11]As before, this tree is schematic and should not be taken to be an endorsement of a rightward-branching analysis of Luvale word order

5 Additional Issues and Problems

5.1 Comparatives

If Kayne (2003, 2005) is correct about the comparative form of *few* modifying NUMBER as well, this analysis provides a straightforward account for phrases such as (40):

(40) fewer than three men

If the underlying structure of (40) was (41), then it would be comparing NUMBER to men:

(41) fewer NUMBER than three men

However, if the underlying structure is really (42):

(42) fewer NUMBER than three NUMBER men

Then one NUMBER is compared to another. However, this cannot be the whole story, as this cannot straightforwardly explain examples such as (43), in which the numeral *a hundred* was argued not to contain NUMBER:

(43) fewer than a hundred men

5.2 Numerals and Syntactic Catagories

A separate question not addressed by the above discussion, but important in many ways, is what does the variation in the syntactic categories of numeral words say about the status of the noun/adjective distinction within linguistic theory. Theories that attempt to derive the noun/adjective distinction on semantic grounds, such as Baker (2003), would, *prima-facie*, find it difficult to explain why certain numerals are nominal and others are adjectival. This is made even more problematic given that different languages draw the line at arbitrarily different points (for a clear demonstration of this fact within the Bantu language family, see Atkins (1961)). A second, equally important, question is what implications does the existence of abstract, unpronounced nominals such as NUMBER have on such theories.

5.3 *One*

One additional open question is how the numeral *one* fits into the pattern described above. In Hebrew, unlike all other numerals, *exad* ('one') follows the head nouns together with the adjectives. In several Bantu languages such as Siswati and Zulu, the numeral that means 1 shows a different agreement pattern than other numerals. It may be the case that *one*, unlike other adjectival numerals, modifies the head noun directly.

If the right-node analysis above is correct, this poses a problem for additive numerals such as *thirty-one*. It may be that in these cases, *one* does modify

NUMBER. This is supported by the fact that in Hebrew, when *one* is conjoined with a higher numeral, it precedes the head noun, unlike normal word ordering:

(44) shloshim ve-axat anashim
thirty and-one people
'thirty-one people'

6 Conclusion

This proposal extends Kayne's (2003, 2005) account of *few* and *many* in a manner that is compatible with Ionin and Matushansky's (2004a, 2004b) observations on the syntax and semantics of numerals. Drawing its support from a variety of cross-linguistic data, some of which has been presented above, it aims to show that despite the fact that low and high numerals seem to belong to different syntactic categories within many languages, numeral phrases are still universally NPs.

References

Ashton, Ethel. O., E. M. K Mulira, E. G. M. Ndawula, and A. N. Tucker. 1954. *A Luganda grammar*. London: Longmarns, Green and Co.

Atkins, Guy. 1961. Notes on the concords and classes of Bantu numerals. *African Language Studies* 2:42–48.

Baker, Mark. 2003. *Lexical categories: Verbs, nouns and adjectives*. Cambridge: Cambridge UP.

Corbett, Greville G. 1978. Universals in the syntax of cardinal numerals. *Lingua* 46:355–368.

Greenberg, Jospeh H. 1978. Generalizations about numeral systems. In *Universals of human language*, ed. Jospeh H. Greenberg, volume 3, 249–297. Stanford, CA: Stanford UP.

Horton, A. E. 1949. *A grammar of Luvale*. Johanesburg: Witwatersrand University Press.

Hurford, James R. 1975. *The linguistic theory of numerals*. London: Cambridge UP.

Hurford, James R. 1987. *Language and number*. Oxford: Basil Blackwell.

Ionin, Tanya, and Ora Matushansky. 2004a. A healthy twelve patients. Paper presented at GURT 2004.

Ionin, Tanya, and Ora Matushansky. 2004b. The non-existence, syntax, and semantics of plurals. USC & CNRS/Universit Paris 8/ENS Ms.

Jespersen, Otto. 1969. *Analytic syntax*. New York: Hok, Rinehart and Winstein.

Kayne, Richard S. 2003. Silent years, silent hours. In *Grammar in focus. Festschrift for Christer Platzack*, ed. Lars-Olof Delsing, Gunlög Josefsson, Halldór Sigurdsson, and Cecilia Falk, volume 2, 209–226. Lund: Wallin and Dalholm.

Kayne, Richard S. 2005. On the syntax of quantity in English. In *Movement and silence*. New York: Oxford UP.

Zadka, Yizhak. 2001. Logic - syntax and meaning. *Hebrew Linguistics* 49:35–55. (in Hebrew).

Zamparelli, Roberto. 2004. Every two days. *Snippets* 9:19–20.
Eytan Zweig, NYU Linguistic Department
719 Broadyway, 4th floor
New York, NY 10003, USA

eytanz@nyu.edu