Anatomy of a complex numeral: Overcounting, with special attention to Ch'ol

Éva Dékány*

Abstract. This paper provides a comprehensive cross-linguistic overview of overcounting, whereby a target numeral is expressed by counting toward the next-higher multiple of the base. I identify three major morphological patterns in overcounting numerals: P-connector, V-connector and no overt connector. I then zoom in on the structure of overcounting numerals in Ch'ol (Mayan). I argue that these numerals are construed with a covert, latively interpreted P, whose silence is due to P-drop.

Keywords. complex numeral; overcounting; Ch'ol; classifier; P-drop

1. Introduction. Based on the semantic relationship that holds between their components, complex numerals can be divided into various subtypes. In multiplicative numerals, the relationship between the components is multiplication, e.g., *five hundred*, *five million*. Additive (aka under-counting) numerals comprise i) the next-*lower* multiple of the base (the augend) and ii) a smaller numeral that is added to this round number (the addend). In the additive numeral *fifty-eight*, for instance, 'fifty' is the augend and 'eight' is the addend. The components of subtractive (aka back-counting) numerals are i) the next-*higher* multiple of the base compared to the value to be expressed (the minuend) and ii) a smaller numeral that is subtracted from this round number (the subtrahend). The Latin expression for 58, for instance, makes reference to 'sixty' rather than 'fifty', and names 'two' as the subtrahend.[1]

* This research has been funded by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences, the ÚNKP-23-5 New National Excellence Program of the Ministry for Innovation and Technology and grant NKFI-145985 of the National Research, Development, and Innovation Fund. I would like to express my heartfelt thanks to the following colleagues for discussion and clarification of data points: Isabelle Bril, Irina Burukina, Anders Holmberg, Carol Rose Little, Brigitte Pakendorf, Alexander Pfaff, Fedor Rozhanskiy, Mária Sipos, Anne Tamm, Nathaniel Torres, Nadya Vinokurova and George Walkden. Ch'ol data not otherwise annotated are from the Tila dialect. The orthography of examples from earlier resources on Ch'ol have been updated and appear in the current standard orthography; I thank Jessica Coon for helping with this. Examples from other languages are presented as spelled in the original. With the exception of recent generative work on Ch'ol, in almost all cases my sources lack a gloss: they provide only the example in the original and a translation (or the numerical value with Arabic numerals). I provide the interlinear glosses myself based on explicit, detailed text-description of the composition of complex numerals in my sources. When this was not available, I confirmed the glosses from grammars or reached out for clarification to colleagues who work on these languages. The usual disclaimers apply. For ease of parsing in English, I glossed the 'twenty' of vigesimal languages as 'score'. For reasons explained in the text, I made an exception for Ch'ol, where I glossed 'twenty' as CL.20. I thank Irina Burukina, Jessica Coon and Rodrigo Ranero for useful feedback on an earlier version of this paper. All errors of fact or fancy are mine. Author: Éva Dékány, HUN-REN Hungarian Research Centre for Linguistics and ELTE Eötvös Loránd University, Budapest (dekany.eva@nytud.hu).

[1] Abbreviations are as follows: A: set A person marking, ABL: ablative, ACC: accusative, ART: article, B: set B person marking, CL: classifier, DEF: definite, GEN: genitive, ILL: illative, IV: status marker for intransitive verb in imperfective, M: masculine, N: neuter, NCL: noun classifer, NML: nominal suffix, NMN: nominalizer, NOM: nominative, ORD: ordinal, PART: partitive case, PFV: perfective, PFX: prefix, PL: plural, POSS: possessive, PREP: preposition, PST: past, RN: relational noun, S: strong, SFX: suffix, SG: singular, W: weak.

(1) Latin
duo-dē-sexāgintā
two-from-sixty
'fifty-eight'

Additive numerals are the cross-linguistically most frequent type, multiplicatives are frequent, and subtractives are rare (Greenberg 1978: 257–258; Hurford 2003: 601).[2]

In addition to the types mentioned so far, there is a further type called overcounting numerals, which occur "sporadically" (Hurford 2003: 601).[3] Similarly to subtractives, overcounting numerals anticipate the next-*higher* multiple of the base compared to the target number. However, instead of applying subtraction, overcounting expresses the target numeral "as so many towards" this higher round number (Comrie 1997: 53). That is, the principle of overcounting is "to express a number as a point on the way to a multiple of the base" (Mazaudon 2008: 4). Compare the overcounting example from Northern Mansi (Uralic) in (2) with its back-counting counterpart in Latin (1). Both languages express 58 with reference to 60, but while Latin counts *backwards* from 60 ('two from sixty'), Northern Mansi counts *in the direction of* 60 ('eight toward sixty').

(2) N. Mansi (Virtanen et al. 2021: 28; original Cyrillic transliterated with copius.eu/trtr.php)
χotpan nupəl ńololow
sixty toward eight
'fifty-eight' (lit. eight toward sixty)

(2) expresses virtual movement along the number sequence. The starting point of this movement is the next-*lower* multiple of the base (here 50), but this reference point normally remains linguistically unexpressed in overcounting numerals. In thought, we move forward from this point along the counting sequence. To express the target numeral, we name i) the next-*higher* multiple of the base (aka the 'prospective'; here 60) and ii) the number of steps we take along the counting sequence towards this higher numeral (here 8). For lack of an established term, I refer to the latter component of overcounting numerals as "the smaller numeral".[4]

---

[2] The inverse of multiplication, namely division, "occurs very rarely" (Hurford 2003: 601). Division, like subtraction and overcounting, makes reference to the next-higher multiple of the base (Comrie 1997). Hurford mentions Welsh *hanner cant* lit. 'half hundred', i.e., 50, as an example, but adds that "It is not clear how well integrated this expression is into the grammar of attributive noun modifiers." Given Greenberg's (1978: 261) observation that "division is always expressed as multiplication by a fraction", we might take division to be a special subcase of multiplication.

[3] The term "overcounting" was coined by Menninger (1969). Greenberg (1978: 258) uses the alternative label "going-on operation", though this term has not gained wide currency in the literature.

[4] Although it is rare, overcounting numerals in some languages have three major components, because in addition to the prospective and the smaller numeral, they also include the next-*lower* multiple of the base in the complex numeral. This is typical of (but not exclusive to) overcounting languages in North-Central Vanuatu (Ochiai 2014). Examples are given below. In the Lhōtā Nāgā example in (4), we can observe overt coordination between the next-lower multiple of the base 'twenty' and the overcounting expression 'one toward thirty'.

(3) West Tamabo (Jauncey 2011: 613)
ngalai-tolu ngalai-vati-na arna
ten-three ten-four-ORD two
'thirty-two' (lit. thirty, fortieth, two)

(4) Lhōtā Nāgā (Witter 1888: 27)
mekwü sü thamdro-we ekhā
twenty and thirty-toward one
'twenty-one'

Most languages that include the next-lower base in overcounting numerals recruit names of body parts to name numerals (Hanke 2005). Hanke (2005) mentions Rawa (Trans-New Guinea) as an example. In this language, 5 is lit.

Multiplicative, additive and subtractive numerals have raised wide interest and much discussion from both typologists and formal linguists (Menninger 1969; Hurford 1975, 2003; Comrie 1997; Greenberg 1978, 2000; He 2015; He et al. 2017; Ionin & Matushansky 2018; Žoha et al. 2022). Perhaps due to its rarity, however, the overall attention that overcounting has attracted has been fairly poor. This paper makes two main contributions. Firstly, it gives the most thorough cross-linguistic survey of overcounting numerals to date, investigating the questions of which languages feature overcounting, for which numerals, and what type of overt morphological connectors are used in these complex numerals (section 2). Secondly, as a nod to Masha Polinsky's contributions to Mayan linguistics, the paper zooms in on the structure of overcounting numerals in Ch'ol and provides the first explicit formal analysis in this domain (section 3).

2. A cross-linguistic survey of overcounting. Based on the morphological connectors (Greenberg's 1978 "links") they feature, overcounting numerals fall into three major types: P-connector (adposition or case), verb-based connector and juxtaposition (i.e., lack of an overt connector). We look at each type in turn. There are also overcounting languages in which I have not been able to identify the category of the overt connector; these are mentioned in section 2.4.

2.1. P-CONNECTOR. Starting with languages with a P-connector, the link typically has the meaning 'toward' (but see below). In this group we have already seen the Uralic language Northern Mansi (2); here overcounting is used for the interval (i.e., non-round) numerals from 21 to 99 (Bakró-Nagy et al. 2022: 546). Overcounting featuring 'toward' also occurs in three Northern varieties of the closely related Khanty language: Honti (1993) gives such forms from Kazym Khanty (only for 'thirty-one' and 'one hundred and one') as well as Shurishkary and Obdorsk Khanty (both 'twenty-one' and upwards).

(7) Obdorsk Khanty (Honti 1993: 296)
χol'-joŋ-pela-kātən
three-ten-toward-two
'twenty-two'(lit. two toward thirty)

The more distantly related Saami group – with the exception of Southern varieties – uses overcounting with the illative case for 11–19; see (8). In the Western varieties and Inari Saami, this pattern is used for higher interval numbers, too (Honti 1993). In the even more distant relative Tundra Nenets (Jamal dialect), similar examples with a lative P can be found for 11–19, as well as for higher interval numerals (Honti 1993).

'hand-one', 10 is lit. 'hand-two', and 15 is lit. 'hand-two leg', see (5). 15 is a base for further compound numerals. An overcounting example is given in (6): 11 is constructed with reference to 15.

(5) Rawa (Toland & Toland 1991: 101)
kande eraya; kande eraya ke-ngga
hand two hand two leg-DEF
'ten'; 'fifteen'

(6) Rawa (Hanke 2005)
kande eraya ke-gidemboro gura-nangge
hand two leg-plus one-only
'eleven'

That (6) involves overcounting is clear; the bracketing internal to this phrase is not fully transparent, though. Hanke's discussion suggests *[kande eraya] [ke-gidemboro gura-nangge]* 'two hands, (and towards the) leg, plus one'. This would involve 'two hands' as the next-lower base, 'leg' as the prospective and 'one' as the smaller numeral. However, to my mind *[kande eraya ke]-[gidemboro gura-nangge]* '(towards) fifteen, plus one' is also a possible internal bracketing here. On this parse, the complex numeral would have just two major parts: the prospective 'fifteen' and the smaller numeral 'one'. This issue will have to be left for further research.

(8) N. Saami (Valijärvi & Kahn 2017: 110)
njeallje-nuppe-lohkái
four-second-ten.ILL
'fourteen' (lit. four into [the] second ten)

(9) Tundra Nenets (Honti 1993: 304)

śidʹi	judʔ	ńaʔ	tʹet
two	ten	toward	four

'fourteen' (lit. four toward twenty)

Overcounting with a 'toward'-connector is also found in some languages of Nagaland, a state in the North Eastern Region of India. The languages mentioned here are all Sino-Tibetan. Lhōtā Nāgā "often" uses this pattern for interval numerals ending in 'one' (4), but this is a less frequently used possibility for other interval numerals above 20 as well (Witter 1888: 27). Nzong (aka Rengma, Southern Rengma or Western Rengma) uses overcounting with 'toward' for all complex numerals ending in 6, 7, 8 and 9; see (10).

(10) Nzong (Mills 1937: 291)

sherü	pamo	tsaro
thirty	towards	six

'twenty-six'

Two further languages in Nagaland with overcounting are Ntenyi (aka Northern Rengma) and Eastern Rengma (aka Anyo, spoken in the villages of Meluri, Sahunyu and Lephori).[5] Ntenyi applies overcounting with 'toward' to all complex numerals ending in 6, 7, 8 and 9; see (11). Eastern Rengma uses overcounting in a number of different morphological patterns; a 'toward' connector appears in 16–19; see (12).[6]

(11) Ntenyi (Mills 1937: 291)

chaa	she	tüo
thirty	towards	six

'twenty-six'

(12) Eastern Rengma (Mills 1937: 291)

mükwe	shun	toro
twenty	towards	six

'sixteen'

Overcounting with 'toward' is also employed in Southern Hokkaido Ainu (isolate, Northern Japan). In this vigesimal language, 40, 60, 80, etc. are formed by multiplication of 20, while 30, 50, 70 and other odd multiples of 10 involve overcounting, as in (13). (See Ochiai 2021; Dékány 2022 for detailed expositions that these are indeed overcounting numerals.)

(13) Southern Hokkaido Ainu (Tamura 1988/2000: 255; glosses based on Ochiai 2021)

wan-pe	e-tu-hot
ten-NMN	toward-two-score

'thirty'

[5] Eastern Rengma split off from Ntenyi in the not-so-distant past (Mills 1937: Part I). While working on his 1937 book, Mills was informed by the Eastern Rengmas that this had happened 16 generations before. Speakers of the two languages "can usually, but not always, understand each other", but they are both mutually unintelligible with Nzong (Mills 1937: 286).

[6] Matisoff (1997) includes two further languages with overcounting from Nagaland: Pochury and Meluri (sometimes written as Maluri). These, however, appear to be the same as Eastern Rengma. Van Driem (2007: 336) explicitly states that Maluri is also known as Eastern Rengma, and this is confirmed in Chukhapa (2021) as well. As for Pochury, Matisoff (1997) and van Driem (2007) treat it as a separate language, and Coupe (2007: 119) describes it as a dialect of southern Sangtam, but according to the Catalogue of Endangered Languages (2024), the names Pochuri, Eastern Rengma, Meluri and Anyo all refer to the same language (see also Chukhapa 2021). The issue of where to draw the line between dialects and languages is known to be a recalcitrant one. For expository purposes, I treat Eastern Rengma, Meluri and Pochury as members of a dialect continuum.

In the Indo-European family, Icelandic can use overcounting with directional marking to express vague cardinalities. The following example has been provided by Alexander Pfaff (p.c.) from a Google search. The basic meaning of the preposition *á* is 'on', but depending on the context, 'in' or 'at' are also possible readings. Like many Ps in German, *á* assigns accusative case in directional expressions and dative in locative expressions. Here it clearly occurs with the former; the literal reading of the clause-initial expression is 'into the second hundred'. The peculiarity of this expression is that the smaller numeral is linguistically unexpressed and semantically vague.

(14) Icelandic

Á	annað	hundrað	látnir	eftir	sprengjuárás.
on	other.N.ACC	hundred.N.ACC	dead	after	bomb.attack.

Hundrað	og	þrír	hið	minnsta	eru	látnir.
hundred	and	three	at	least	are	dead

'A cardinality in the second hundred (i.e., more than 100 people) dead after a bomb attack. At least 103 are dead.'

18th century Yucatec Maya also used a P-connector in its overcounting numerals (see Beltran 1746 [1859]: 195–201 for the forms and Yasugi 1995 for their segmentation). The P used in these cases is *t*, which corresponds to (a contracted form of) the all-purpose preposition of modern Yucatec Maya *ti'* (covering the functions of English 'to, of, at, on, in, for, by'; Andrade 1955: Ch. 2.29.). As Colonial Yucatec Maya is vigesimal, the next-higher 'round' numeral is always a multiple of 20. Thus 43, for instance, is constructed with reference to 60, rather than 50.[7]

(15) 18th c. Yucatec Maya (Yasugi 1995: 307)

ox	t-uy-ox-kal	
three	PREP-A3-three-score	

'forty-three'

So far we have seen the following types of P-connectors: the all-purpose preposition of languages with just one P, a 'toward' P and the illative case. Additionally, the genitive case may also serve as a linker between the prospective and the smaller numeral. In Old Norse (Indo-European), the examples I have seen involve a discontinuous numeral (see also section 3.1), with the noun wedged between the smaller numeral and the prospective, as in (16).

[7] In Beltran's list, overcounting is used systematically, starting with 21, and *t* appears on the vast majority of these numerals. The exceptions are i) 171, 370 and 390; and ii) when the smaller numeral is 10 or 15 and the overall value of the complex numeral is below 175. In these cases, the smaller numeral and the prospective are juxtaposed without a connector. Yasugi (1995: 308) remarks on this in the following way: "It is difficult to explain this irregularity. It may be due to carelessness, or it may reflect a quinary method in use, although the numbers in the interval are formed on a decimal method." Tozzer (1921: 103), on the other hand, suggests that the systematic lack with the 10 and 15 "units seems to show some definite purpose when it is omitted." Although it is difficult to say what this purpose could be, it is worth noting that the 10 and 15 units are not entirely random points on the number sequence of a vigesimal language. The 10 unit corresponds to the halfway point between two multiples of the base, and from here, the 15 unit is halfway through the remaining distance to the prospective. The 10 and 15 units are special in Dzongkha (Sino-Tibetan) as well. In the vigesimal system of this language, it is precisely the 10 and 15 units between two multiples of the base (i.e., 30, 35, 50, 55, 70, 75, etc.) that are expressed via overcounting; see the Appendix. Colonial sources after Beltran and modern grammars only report additive numerals. Overcounting numerals were replaced with additives in the history of some other Mayan languages as well, e.g., Kaqchikel, K'iche' and Mam.

(16) Old Norse (Menninger 1969: 76; glosses courtesy of George Walkden p.c.)
II menn hin-s ellift-a tig-ar
two man_PL ART-GEN.SG eleventh-M.W.GEN.SG ten-M.GEN.SG
'102 men' (lit. two men of the eleventh ten)

In Lithuanian (Indo-European), fractions with 'half' comprise the word 'half' and the ordinal form of the next-higher integer in the genitive (17). Menninger (1969) calls this pattern the "half-count". Similar fractions, limited to time-telling expressions, are used in Russian as well.

(17) Lithuanian (Ambrazas 1997: 173)
pus-trẽceio, pus-šẽšto
half-third.GEN half-sixth.GEN
'two and a half', 'five and a half'

(18) Russian (Irina Burukina p.c.)
pol-pjatogo
half-fifth.M.SG.GEN
'half past four'/'4:30'

In addition to the P-connectors seen so far, the ablative case has been implicated in overcounting in two Siberian languages, Evenki (Tungusic) and Sakha (aka Yakut, Turkic) (Pritsak 1955). Since the meaning of the ablative is exactly the opposite of that of the lative 'toward', its use in overcounting might seem surprising at first, especially in light of the fact that the ablative is also used in *subtractives* (1). How could one language use 'toward', and the other 'from', to construct an overcounting numeral?

As pointed out by Alexander Pfaff (p.c.), the formally possessive phrase in (16) is likely best interpreted as a partitive expression ('two men out of the eleventh ten'). A partitive interpretation is plausible for (17) as well ('half out of the third'), since the Lithuanian genitive has partitive uses (Seržant 2014). In some languages, partitive readings are coded with the ablative rather than the genitive. The ablative–partitive connection is evident in the history of Finnic, too: here the morphologically distinct partitive case developed out of the ablative (Grünthal 2022). Consequently, if a partitive reading were available for the ablative of Evenki and Sakha, then we could make sense of the use of this case with overcounting numerals.

That said, there is some doubt as to whether the relevant data in Evenki and Sakha even involve numerals proper. In Sakha, the forms in question indicate one's age in a possession sentence. Pritsak gives the example *sättä uommuttan ikkitin ïllïm* and translates it as "von meinen 70 [Jahren] habe ich 2 genommen" (lit. 'I took 2 of my 70 [years]', i.e., 'I am 62'). (19) shows the morphological decomposition of this example without the clause-final verb *ïllïm* 'take.PST.1SG' (glosses from Nadya Vinokurova p.c.).

(19) Sakha (Pritsak 1955: 191)
sättä uom-m-uttan ikki-tin
seven ten-1SG-ABL two-3SG.ACC
'(my) sixty-two' (lit. from my 70, 2)

(20) Sakha: suggested analysis
(min) sättä uom ~~saas~~-m-uttan ikki
(I) seven ten ~~age~~-1SG-ABL two
'from my seventy [years], two'

Importantly, the prospective in (19) is marked not just for ablative, but also for 1SG. This indicates that we are looking at a possessive construction. The possessive NP in Sakha involves a morphologically unmarked possessor, which can undergo *pro*-drop, and a possessum that agrees with the ϕ-features of the possessor. The 1SG marking points to a *pro*-dropped 1SG possessor: (19) is literally 'from my 70, two'. The possessed noun is understood to be *saas* (lit. 'spring/age', i.e., 'years'), but this noun is not pronounced. I suggest that it has been elided, and its suffixes, the 1SG possessive agreement and the ablative case, have been stranded under ellipsis. These now lean onto the adjacent numeral for phonological support, as in (20). This is a case of so-

called suffix-stranding NP ellipsis, a phenomenon discussed for agglutinating languages in Lipták & Saab (2016), and which Nadya Vinokurova (p.c.) confirms is a feature of Sakha syntax.

Although the logic in (19) is indeed that of overcounting (with 70 being the prospective and 2 the smaller number), this expression can hardly be called a complex *numeral*. Compare English *Of my ten reindeer, two have died*, where *of my ten reindeer, two* is a complex expression but not a complex *numeral*. It is unclear from Pritsak's text if an expression like (19) can also occur outside of possession sentences, in garden-variety adnominal position to non-possessed nouns (thus if 'sixty-two reindeer died', for instance, could be expressed as 'seventy-ABL two reindeer died'). We should therefore be cautious in positing the existence of overcounting with an ablative P.

Turning to the purported ablative connector in Evenki, here the situation is even less clear. Citing Vasilevič (1958), Pritsak gives *d'ūr-d'ā-kin d'ūr* as a dialectal form for 12. He identifies *d'ūr* as 'two', *d'ā(n)* as 'ten', and concludes that *-kin* "must be" the ablative case ("kann es sich nur um einen Elative/Ablativ handeln"; Pritsak 1955: 191). Pritsak bases this on the form of the ablative in other North Tungusic languages, not on his knowledge of, or sources on, Evenki itself. He further indicates that *-kin* should be decomposed as *-ki-n*. With *-n* being the 3SG possessive suffix (Nedjalkov 1997: 143), the question arises whether the Evenki examples are also elliptical possessive NPs rather than real complex numerals. In any case, I could not confirm the use of *-kin* (or *-ki*) as an ablative from Evenki grammars; Däbritz & Gusev (2021) gloss *-kin* as a nominalizer.

2.2. V-BASED CONNECTOR. Overcounting with a verb-based connector could be found up until the 1920s in varieties of Ao (Nagaland, Sino-Tibetan). In Mongsen, Changki and Longla Ao, this definitely applied to 16–19 (and perhaps also to some higher interval numerals) (Mills 1926: 342), while in Chungli Ao there is evidence for overcounting in the higher compound numerals ending in 6–9 as well (Avery 1886; Clark 1893). In all of these Ao dialects, the morphological connector is a negated verb. Coupe's (2007) contemporary Mongsen consultants explain this way of constructing the number 19 as "(the) twenty not completed, (the) nine" (Coupe 2007: 118).

(21) Mongsen Ao (Mills 1926: 342)
mükyi mü-pen tükū
twenty not-complete nine
'nineteen'

(22) Chungli Ao (Clark 1893: 45)
semu̯r ma-ben trok
thirty not-bring six
'twenty-six'

Two further Sino-Tibetan languages of Nagaland, Angami and Sema, also use overcounting with an overt connector. In Angami this applies to all interval numerals ending in 7, 8 or 9 (McCabe 1887: 15). In Sema, overcounting is obligatory for complex numerals ending in 9 and optional for those ending in 6, 7 or 8 (Hutton 1916: 3–4). The connectors are not identified in the grammars cited above, but Coupe (2012: 210) makes it explicit that these are also negated verb stems.[8]

[8] Coupe (2012: 210) also claims that the connector *pamo* of Nzong is a negated verb stem; see (10). Although the clausal negator of Nzong is indeed the *-mo* verbal suffix, and *pa* is somewhat similar to the Mongsen Ao verb *pen* 'complete', I have based my classification on Mills's (1937: 291) grammar, which includes *pamo* among its list of "suffixes" with the meaning 'toward'. Ntenyi and Eastern Rengma – which are related closely enough to be investigated alongside Nzong in the same monograph by Mills – also use a 'toward' connector. (In these languages, there is no segmental overlap between the verb negator and the connector of overcounting numerals.) That said, it is possible that in Nzong there is a diachronic or synchronic connection between 'toward' and a negated verb such as 'not complete, not reached'.

2.3. JUXTAPOSITION. Some languages forgo an overt connector between the prospective and the smaller numeral. In the Turkic family, Old Turkic used overcounting without a morphological connector for the interval numbers between 11 and 99 (Ehlers 1983; Clark 1996).[9] The only modern Turkic language that employs overcounting is Sarig Yugur (Sunan-Yugur Autonomous County of Gansu province in China), where such compound numerals are used for the interval numbers between 11 and 29 (Clark 1996).

(23) Old Turkic (Erdal 2004: 220)
tört kïrk; säkiz tokuz on
four forty eight nine ten
'thirty-four'; 'eighty-eight'

(24) Sarig Yugur (Clark 1996: 27)
per yïGïrma; pis otïs
one twenty five thirty
'eleven'; 'twenty-five'

There are also a good number of Mayan languages that use overcounting. Examples include Classical K'iche', Classical Kaqchikel and Classical Mam (all from 41 up), Classical Chontal (from 21 upwards), as well as Tsotsil, Tseltal, Popti' and Chuj (all starting with 21), Ixil (from 41 up) and Ch'ol (see section 3).[10] Apart from the lack of a morphological connector, these numerals are built similarly to the Colonial Yucatec Maya example given above; see (25) for Classical Kaqchikel. (Section 3.2 contains detailed argumentation why I do not find it plausible that the Set A morphology is a linker between the prospective and the smaller numeral.) Overcounting without a morphological connector also occurs in varieties of the geographically closely located Zapotec group (Oto-Manguean): Classical Zapotec (starting with 41), as well as Yatzachi and Juárez Zapotec (between 21 and 59); see (26).

(25) Classical Kaqchikel (Yasugi 1995: 105)
xun ru-xu-muč'
one A3-one-eighty
'sixty-one'

(26) Yatzachi Zapotec (Yasugi 1995: 285)
žda'-ayon
fourteen-sixty
'fifty-four'

In the Uralic family, overcounting without a morphological connector is used in Central Mansi dialects for 18 and 19 (Honti 1993). A number of Khanty varieties have this, too: Demyanka Khanty (for 18 and 19), Vakh-Vasyugan and Surgut Khanty (for 80), as well as Salym, Kazym, Obdorsk, Central Ob and Serkali Khanty (for 18 and 80); see Honti (1993).

(27) Central Mansi (Honti 1993: 298)
mōt-low-ńallow
other-ten-eight
'eighteen' ('other' = 'second')

[9] Although there is no overt connector that could indicate the nature of the relationship between the two parts of the numeral, and native speakers are not available to confirm the meaning of these numerals either, there is consensus among Turkologists that these numerals involve overcounting rather than addition or subtraction. On the interesting history of how this consensus has been reached, see He (2022).

[10] According to Yasugi (1995: 104), overcounting is absent only from Tojol-ab'al, Tz'utujil, and modern Kaqchikel, but this is inaccurate. First, Brown et al. (2006: 27) note that overcounting is still used in some rural Kaqchikel communities, and second, overcounting does not appear to be used in Modern Mam and Modern Yucatec Maya either.

The construction of ‘eighty’ is somewhat unexpected in these dialects, as it comprises ‘eight’ juxtaposed to ‘hundred’; literally ‘eight-hundred’; see (28). These are arguably to be understood as ‘eight [tens toward a] hundred’, with ‘ten’ remaining covert. (Note that ‘ninety’ is formed in the same way in Russian as well.)[11]

(28) Kazym Khanty (Honti 1993: 295)
ńĭwəʌ-sɔt
eight-hundred
‘eighty’

In the same family, the Finnic languages (Finnish, Estonian, Ingrian, Votic, etc.) currently use overcounting without a morphological connector in the 11–19 range, but until the end of the 19th century, overcounting applied to the higher numerals as well (Honti 1993). In overcounting numerals, the smaller numeral is followed by the expression ‘second ten’. However, the ‘ten’ is normally dropped, and thus the normal formula for these numerals is ‘smaller numeral-second’ (e.g., ‘eleven’ is literally ‘one second’, meaning ‘one in/toward the second ten’). Illustrative examples are provided below.[12]

(29) Ingrian (Markus & Rozhanskiy 2022a: 316)
ükš-toišt(-kümmend)
one-second.PART (-ten.PART)
‘eleven’

(30) Votic (Markus & Rozhanskiy 2022b: 338)
ühs-tõjššümed
one-two.PART.ten
‘eleven’

In the Formosan group of the Austronesian family, similar forms are used mostly by older speakers of the Northern Amis dialect: here, too, the smaller numeral is juxtaposed to the ordinal form of the multiplier of the next-higher ten, but the multiplicand ‘ten’ itself remains covert. The range of overcounting numerals in this language is not known.

(31) Northern Amis (Isabelle Bril’s fieldwork data, obtained via p.c.)
lima saka-pitu
five ORD-seven
‘sixty-five’ (lit. five [in/toward the] seventh [ten])

Among the languages of Nagaland, Eastern Rengma uses overcounting without a morphological connector for all interval numbers starting from 21. (We have seen that this language uses a ‘toward’ connector for 16–19; cf. (12)). In the 21–89 range, the next-higher multiple of the

[11] The numerals 19 and 90 in these Khanty dialects should be looked into further. Honti (1993) argues that originally they were subtractives, but over time the morphemes meaning ‘one’ and ‘lacking/missing’ have undergone phonological reduction and fusion, to the extent that the original meaning has become opaque. Thus synchronically they may be overcounting numerals.

[12] In present-day Finnish, ‘ten’ must be dropped, while Livonian does not allow this type of truncation (Honti 1993). The partitive case in these Finnic numerals is the “partitive of quantification”, whose distribution is similar to that of the “genitive of quantification” in Russian. It is not specific to overcounting but occurs internally to multiplicative numerals as well.

base bears the *a-* prefix (32), while in the 91–99 interval an additional *-wa* suffix appears, too (33).

(32) Eastern Rengma (Mills 1937: 291)
a-chera kesü
PFX-thirty one
'twenty-one'

(33) Eastern Rengma (Mills 1937: 291)
a-meza-wa kesü
PFX-hundred-SFX one
'ninety-one'

Mills (1937) notes that *a-* occurs on almost all nouns, and it can also be observed that many locative words, including 'among, outside, above, by the side, into', also begin with this prefix (these may be locative nouns). Mills makes no remarks on the *a-* of numerals; it may or may not be identical to the *a-* of nouns. In any case, it does not appear to function as a morphological linker between the two parts of the complex numeral. The *a-* prefix and the *-wa* suffix together constitute the ordinal form of numerals (cf. *keche* 'three', *a-keche-wa* 'third'). These numerals thus may be partially similar to 11–19 in Finnic, which also involve an ordinal.[13] Tamabo (Austronesian), already exemplified in (3), uses overcounting for all interval numerals. Similarly to Finnic, the smaller numeral is juxtaposed to the ordinal form of the prospective (Jauncey 2011).

Overcounting also occurs in the Indo-European family, with fractions expressed with the half-count (see Lithuanian (17)). This can be observed in (earlier) German and Classical Danish (Garczyński 2014), as well as the modern Scandinavian languages (Alexander Pfaff p.c.). Diachronically, complex numerals with this logic were not restricted to fractions. This is shown in (35) from Old Norse, where the 'halved' item is not an integer but a specific set of tens (glosses from Alexander Pfaff).

(34) Cl. Danish (Garczyński 2014: 30)
halv-tredje
half-third
'two and a half'

(35) Old Norse (Menninger 1969: 78)
half-uhr tiund-e tugh-r
half-S.M tenth-W.M ten-NOM.SG
'ninety-five' (half of the tenth ten)

Fractions of the type in (34) also occur as components of 50, 70 and 90 in Classical Danish and Faroese (37).[14]

(36) Cl. Danish (Garczyński 2014: 30)
tre-sinds-tyve
three-times-score
'sixty' (three scores)

(37) Cl. Danish (Garczyński 2014: 30)
halv-tred-sinds-tyve
half-third-times-score
'fifty'

The half-count also occurs in some Uralic languages. In earlier Hungarian, this applied to fractions ending in 'half' (38). Although these are mostly obsolete in the modern language, 'one and a half' is still in active use and can also function as a multiplier of powers of 10 (39).

[13] *-Wa* is also used independently of the *a-* prefix, cf. *aküza* 'small', *aküza-wa* 'the small one', and in temporal expressions it means 'in' (Mills 1937: 292–292).

[14] Although the half-count is regularly treated as a type of overcounting (e.g., Comrie 1997), it should be noted that it may, in principle, also be analyzed as subtraction: 'half-third' (2½) could just as easily be 'half-from-third' as 'half-into/toward-third'. A weak counterargument against the subtractive analysis could be Greenberg's (1978) Generalization 11, which states that the subtrahend and the minuend are always connected by an overt morphological connector. The only exception to this appears to be 'nine' in Dravidian languages ('one, ten'). Additionally, the half-count could also potentially be viewed as multiplication: 'half times the third one'. See also Mazaudon (2008: fn. 7).

(38) Hungarian (own knowledge)
más-fél, ötöd-fél
other/second-half fifth-half
'one and a half', 'four and a half'

(39) Hungarian (own knowledge)
más-fél millió
other/second-half million
'one and a half million'

In earlier Estonian, tens and hundreds could be 'halved' in a fashion similar to (35). Honti (1993: 69) gives a parallel for (40) from an unidentified Saami variety as well.

(40) earlier Estonian (Honti 1993: 69)
pool-kolmat-kümmend
half-third-ten
'twenty-five'

(41) earlier Estonian (Honti 1993: 69)
pool-teist-sadda
half-second-hundred
'one hundred and fifty'

2.4. INTERIM SUMMARY. Cross-linguistically, overcounting may lack an overt connector between the smaller numeral and the prospective. If a morphological linker appears, it tends to be a P-element with the meaning 'toward', though occasionally the genitive case is used as a linker too, and there are also cases of a V-based connector. Many overcounting languages use the ordinal form of the prospective, e.g., 'second ten' instead of 'twenty'. Among these languages, the multiplied base 'ten' remains implicit in a number of cases (e.g., Northern Amis or Finnish). Even though overcounting numerals are often paraphrased in English as '*smaller numeral* in the second ten', I have not found a stative locative P connector ('in') in any language. A P connector with the meaning of 'before' or 'in front of' would also be semantically appropriate (e.g., 'the two before/in front of twenty' for 12), but I have not found examples of this type either. Some languages switch between the no connector and the overt connector strategies depending on the numeral value to be expressed (e.g., Eastern Rengma, Obdorsk and Kazym Khanty).

I have identified two related, interesting cases where the smaller numeral is followed not by the next-higher multiple of the base, but by a higher round numeral. Both occur in Khanty dialects and concern the numeral 101, which is expressed as 'one towards the second hundred' rather than 'one towards one hundred and ten'/'one towards the eleventh ten'. This must be connected to the fact that with 101, the first one-hundred threshold has just been crossed.[15]

(42) Kazym Khanty (Honti 1993: 295)
kĭmət sɔt pelʌĭ ĭt
second hundred towards one
'one hundred and one'

(43) Obdorsk Khanty (Honti 1993: 223)
kimət sat-əl it
second hundred-POSS.3SG one
'one hundred and one'

Overcounting languages exhibit a geographical skewing: they form clusters in Mesoamerica, Northern Europe, parts of Siberia, Nagaland, parts of the Philippines, North-Central Vanuatu and Taiwan. In all cases I have seen, overcounting is used for only a subset of the interval numerals (the others being formed with addition), and sometimes additive synonyms exist for the overcounting numerals themselves (Ch'ol, some Khanty dialects, Lule Saami, among others). The closer the target numeral is to the next-higher multiple of the base in the number sequence, the more likely it is for overcounting to occur. That is, interval numerals ending in 6 through 9 are

[15] In the Obdorsk example, 'second hundred' bears possessive agreement. This does not signal a possessive relationship between 'the second hundred' and 'one', as in 'one of the second hundred'. This is because the Khanty possessive agreement is borne by the possessed noun, and in 'one of the second hundred', the possessed noun would be the *smaller* numeral. In addition to its possessive use, the Khanty POSS.3SG also has a so-called determining use, in which it does not signal possession but functions more like the definite article of article languages (Simonenko 2014). Honti (1993) is explicit that in this example, POSS.3SG has this determining function.

more frequently expressed by overcounting than those ending in 1 through 5, and complex numerals ending in 9 have the highest likelihood of being formed with this method.

There are also a number of Austronesian languages that employ overcounting and that I have not had the opportunity to investigate with an eye toward the typology above, or where I have not been able to identify the categorial status of the connector. The North-Central Vanuatu languages Tutuba, Tangoa, Akei, Big Nambas and one of the Ambae varieties use overcounting above 20. They are similar to Tamabo (also spoken in this area; see (3)) in that they overtly express the next-lower multiple of the base, and place the next-higher multiple of the base in the ordinal form (e.g., 'twenty, thirtieth, one' for 21; see Ochiai 2014). Ochiai (2014) also lists a number of Indigenous languages in Taiwan (e.g., Central Amis, Yami and Paiwan) and some languages in the Philippines (all Austronesian, e.g., Ilokano and Tagalog) that use overcounting. In almost all of Ochiai's examples from Taiwan and the Philippines, there is at least one morpheme in the original that remains unidentified in the interlinear gloss. For instance, 21 in Tagalog is *ma-ika-tlo-ng isa*. This contains 'thirty' and 'one' as its components, but the exact decomposition is opaque because it receives the gloss 'MA-IKA-three-NG one'. This prevented meaningful investigation into what type of connector these languages involve (if any).

There are also Sino-Tibetan languages that need to be looked into further. Dzongkha, Bumthang, Kaike, Gongar, Sharchokpa and Lepcha use the half-count for the odd multiples of 10 (i.e., 30, 50, 70, etc.; see Mazaudon 2008). With the exception of Lepcha, they all contain an overt connector that I have not been able to identify. He (2022) also mentions that overcounting can be found in some Dravidian and Niger-Congo languages, without giving further specifics. In the latter group, Yorùbá may be a case in point (Ẹkundayọ 1977), though a more precise morphemic decomposition is required than the sources available to me allowed. A detailed look at these cases may require a refinement or expansion of the typology given above.

3. Overcounting in Ch'ol. In this section, I zoom in on traditional overcounting numerals in Ch'ol (Mayan) and analyze their internal structure in detail. Like some of its relatives in the Mayan family, Ch'ol is a numeral classifier language. NPs with a native numeral always contain a classifier, which is suffixed to the numeral, as in (44) and (45).

(44) Ch'ol (Bale et al. 2019)
ux-tyikil x'ixik
three-CL woman
'three women'

(45) Ch'ol (Bale et al. 2019)
ux-kojty wakax
three-CL cow
'three cows'

Ch'ol is a vigesimal language: *-k'al* 'twenty' forms the basis of multiplicative numerals. *-K'al* has the distribution of classifiers. It occupies the post-numeral CL position, with a multiplier in the numeral slot (46), and it blocks regular classifiers used with simplex numerals (47).

(46) Ch'ol (Bale et al. 2019)
ux-k'al wakax
three-CL.20 cow
'sixty cows'

(47) Ch'ol (Bale et al. 2019)
*ux-k'al-kojty wakax
three-CL.20-CL cow
Intended: 'sixty cows'

Nowadays native numerals are used only for 1–6, 10, 20, 40, 60, 80, 100 and 400. The other numbers are named with Spanish borrowings. These do not occur with classifiers (Vázquez Álvarez 2011: 160; Bale et al. 2019).

Traditionally, Ch'ol used overcounting for the interval numerals. Although these are not in active use any more, they have been well documented in grammars, and are built in the following

way. The smaller numeral takes a classifier that is appropriate for the counted noun. (48) features *-p'ejl*, which is used for spherical objects and is also the general inanimate classifier, while (49) involves *-tyikil*, the classifier for humans.[16] The smaller numeral and its classifier are followed by the prospective; this has the multiplier in the numeral slot and the base *-k'al* 'CL.20' in the classifier slot. This part of the complex numeral is prefixed with Set A morphology in 3rd person.

(48) Ch'ol (Wycliffe & Bíblica 2011: 1 Reyes 15:33)
chäm-p'ejl i-cha'-k'al jab'
four-CL A3-two-CL.20 year
'twenty-four years' (lit. four two scores year)

(49) Ch'ol (Wycliffe & Bíblica 2010: Apocalipsis 19:4)
chäñ-tyikil i-cha'-k'al ancianoj-ob
four-CL A3-two-CL.20 elder-PL
'twenty-four elders' (lit. four two scores elders)

In Warkentin & Scott (1980: Appendix A) all non-round numerals above 20 involve overcounting. In Merrifield (1966), on the other hand, in the 20–40 interval only 25 and 30–39 are constructed this way; 21–24 and 26–29 are additive. The higher intervals follow the same pattern.

Starting with 400, numerals are based on the powers of 20: *-bajk'* for 400 and *-pik* for 8000. Similarly to *-k'al*, these have the distribution of classifiers. (50) shows a larger numeral that employs the overcounting pattern recursively.

(50) Ch'ol (Wycliffe & Bíblica 2010: Hechos 13:20)
lujum-p'ejl iy-ux-k'al i-cha'-bajk' jab'
ten-CL A3-three-CL.20 A3-two-CL.400 year
'450 years' (lit. fifty eight hundred year)

(48)–(50) raise three questions. i) What sort of phrases are joined in these overcounting expressions? Are these constituents that comprise only the numeral and the classifier ([Num-Cl]-[Num-Cl] N), or do the joined phrases correspond to full extended NPs, each with its own noun under the surface? ii) Why does the A3 morphology appear? iii) How are the two parts of the complex numeral joined syntactically? We look at each question in turn.[17]

3.1. THE SIZE OF THE CONNECTED CONSTITUENTS. There are two major approaches to the structure of complex numerals on the market. On the first one, the parts of the complex numeral form a constituent to the exclusion of the noun: [[Num+Num] N] (He 2015). On the second analysis, each component of the complex numeral independently combines with an instance of the noun: [[Num N]+[Num N]] (Ionin & Matushansky 2018). On this approach, language-specific rules determine if all instances of N are pronounced or some undergo ellipsis under identity.

[16] Human-denoting nouns can, but do not have to, be plural-marked in order to obtain a plural interpretation. Other nouns do not combine with the plural suffix.

[17] When it comes to classifier expressions, there is, of course, also the perennial question of whether the classifier forms a constituent with the noun or with the numeral. Specifically for Ch'ol, this has been discussed in Bale et al. (2019) and Dékány (2024). Question i) does not commit us to a taking a stand on [Num [Cl N]] vs. [[Num Cl] N] if the related phrases turn out to be full NumPs, both with their own numeral, classifier and (potentially elided) noun. Questions ii) and iii) are entirely independent of this issue.

It is logically possible that cross-linguistically (and perhaps even intra-linguistically), both structures are available for building complex numerals (Žoha et al. 2022). However, it would make for a stronger hypothesis – and a simpler UG – if only one of these options were available, and this would simplify the language learner's task, too. It is therefore desirable to try and work with one structure, getting as much mileage out of it as possible. It seems to me that if numeral structures are to be unified cross-linguistically, Ionin & Matushansky's (2018) approach is the way to go. This is because in some languages the noun appears *inside* the complex numeral, as in the Old Norse (16), or Scottish Gaelic *trì fir dheug*, lit. 'three men teen', i.e., 'thirteen men' (Hurford 2003: 598). There are also languages in which an instance of the noun is pronounced after each component of the complex numeral (e.g., Biblical Welsh, see Hurford 1975: Ch. 6.11, or the Bantu language Luvale, see Ionin & Matushansky 2018: 122). This occurs not only in additives, but also in overcounting numerals (Dékány 2022). Such examples can only be captured in Ionin & Matushansky's (2018) analysis. Languages in which the noun occurs only once, after the complex numeral, can be derived in both Ionin & Matushansky's (2018) analysis and He's (2015) approach. The former theory therefore has greater coverage, and I adopt it in what follows.

Ch'ol overcounting numerals contain two classifiers: a regular classifier after the smaller numeral and *-k'al* in the prospective ((48)–(49)). Of particular interest here is the first one, located inside the complex numeral. The mere fact that a classifier appears here does not mean much in and of itself. Ch'ol numerals are morphologically bound, so they do not appear without a classifier, even in the counting sequence (see Little et al. 2022; Dékány 2024 for discussion). What is significant, however, is that this classifier covaries with the noun (in terms of animacy, shape, etc.). I take the "matching" between classifiers and nouns to be a case of selection under locality. If the components of the complex numeral formed a constituent to the exclusion of the noun, as in [[Num-CL A3-Num-CL.20] N], then the first classifier would be deeply embedded inside a modifier, without being in a local relationship to the noun. It is difficult to see how their covariation could be captured in this scenario.[18] Adopting Ionin & Matushansky's (2018) theory of complex numerals to Ch'ol overcounting, we do not run into this problem. The smaller numeral is in a local relationship with an instance of the noun that has been elided (a case of PF-deletion under identity); their covariation is captured as selection under locality. This is sketched in (51).

(51) [chäm-p'ejl ~~jab'~~] [i-cha'-k'al jab']
four-CL year A3-two-CL.20 year
'twenty-four years'

NP/nP ellipsis after a classifier is independently attested in Ch'ol, see Vázquez Álvarez (2011: 157) and Coon (2017: 666). The viability of (51) for Ch'ol is further discussed in Dékány (2024).[19]

[18] The covariation in question is especially interesting when contrasted with data from Tseltal (Mayan). Tseltal overcounting numerals are built similarly to those in Ch'ol, but the smaller numeral must consistently bear the *-Vb* suffix; a specific classifier that semantically "matches" the counted noun is not possible. There is some controversy over whether *-Vb* is a general/default classifier (de León Pasquel 1988: 62 and Kaufman 1971: 92 are in favor, while Fleck 1981: 10 and Polian 2013: 595 are against this). In any case, this pattern shows that just because the numeral is a bound morpheme and so some lexeme is required to phonologically support it, covariation between the smaller numeral and the noun will not be automatic.

[19] It has sometimes been suggested that the semantic "matching" between classifier and noun should rather be captured via a presupposition that the classifier introduces on the noun. If this was the case, then the kind of local

3.2. A3 MORPHOLOGY. Let us turn to the role of the A3 marker. In Mayan linguistics 'Set A' and 'Set B' are terms for two series of person/number markers that cross-reference arguments on a head. In Ch'ol, 'Set A' markers consistently cross-reference an external argument: the ergative argument on a verb, the possessor on the possessed noun and the Ground on a relational noun (Coon 2013: 45). Given that numerals are in the extended projection of nouns, it is natural to take the A3 morphology to reflect some sort of possessive structure in overcounting numerals.

The most straightforward assumption would be that the possessive relationship holds between the smaller numeral and the prospective. Possession does occasionally occur in the English paraphrases of the numerals in question, e.g., 22 as 'two of the group of two-20s' in Bale et al. (2019) or 25 as 'five (units) of the second score' in Merrifield (1966). On the intuition expressed in these paraphrases, the smaller numeral would be the possessee and the prospective would correspond to the possessor. That this is plausible is confirmed by Old Norse (16), where the prospective is marked in the same way as possessors, namely by genitive case. This approach would also immediately answer the question of how the two parts of the complex numeral are linked grammatically: the answer would be 'via a possessive structure'.

However attractive this approach may seem, it cannot be squared with the morphosyntax of Ch'ol overcounting numerals. Consider a garden variety possessive construction. As shown in (52), the possessee precedes the possessor, and Set A morphology appears on the possessee.

(52) Ch'ol (Vázquez Álvarez 2011: 76)

y-ijñam	aj-Wañ	Xañtyes
A3-wife	NCL-Juan	Sanchez

'Juan Sanchez's wife'

(53) Ch'ol

chäm-p'ejl	i-cha'-k'al
four-CL	A3-two-CL.20 year

'twenty-four'

Compare now the numeral 24 from (48), repeated in (53). If 24 were 'the four of the second score/two-20s', then the A3 marker should be prefixed to the smaller numeral, but instead, it must occur on the prospective. From the perspective of the paraphrases above, this is inexplicable.

We could also turn the analysis around, and take the position of the A3 morphology to indicate that the prospective is the possessee, with the smaller numeral being the possessor. This approach has its own problems. From a cross-linguistic perspective, there are cases such that grammatically a possessive relationship holds between the components of an overcounting numeral. However, in these cases the possessor is always the prospective and never the smaller numeral ((16)–(18)). In addition, there are challenges from a Ch'ol internal perspective as well. Syntactically, Ch'ol possessors are strictly postnominal (only *wh*-possessors constitute an exception; Coon 2009). If the smaller numeral were the possessor, it would then be in the wrong position. Semantically, this structure would have the meaning 'the second score/two-twenties of four'. This does not seem to correspond to what the numeral 24 means, though. Based on the foregoing discussion, I conclude that although Ch'ol overcounting numerals do involve

relationship between the classifier and noun that selection requires may not be necessary, and [[Num-CL A3-Num-CL.20] N] may be feasible. There are doubts, however, as to the viability of the presupposition approach. In several languages, the noun classes carved out by sortal classifiers are rather heterogeneous. Although this is not typical in Ch'ol, there are some examples. For instance, the classifier *-kojty* is derived from *-koty*, a positional root meaning 'standing on four legs'. While its use with four-legged animals is not surprising, *-kojty* also counts two-legged and legless animals, as well as chili peppers irrespective of shape (Bale et al. 2019). I challenge the proponents of the presupposition approach to suggest a presupposition that carves out all and only the nouns taking *-kojty*.

possession, the possessive relationship does not take the smaller numeral and the prospective as its two terms.

What *are* the two terms of the possessive relationship, then? I take the A3 morphology on the prospective to indicate that it is one of the terms; specifically, it is the possessee. Thus in (48) and (49), the possessee is 'two-CL.20'. I have argued that the other term, i.e., the possessor, is not the smaller numeral. I propose that the possessor is the overt noun that follows the prospective: 'year' in (48) and 'elder' in (49).[20]

What *are* the two terms of the possessive relationship, then? I take the A3 morphology on the prospective to indicate that it is one of the terms, specifically, it is the possessee. Thus in (48) and (49) the possessee is 'two-CL.20'. I have argued that the other term, i.e., the possessor, is not the smaller numeral. I propose that the possessor is the overt noun that follows the prospective: 'year' in (48) and 'elder' in (49).

(55) [chäm-p'ejl ~~jab'~~] [i-cha'-k'al jab']
four-CL year A3-two-CL.20 year
'twenty-four years' (four years in/of/toward [the] two-twenty [group] of years)

I suggest that in overcounting numerals *-k'al* 'CL.20' does not refer to the number 20 directly, but is rather interpreted as a group or set of numbers: *cha'-k'al* 'two-CL.20' is the set of numerals between 21 and 40, *ux'-k'al* 'three-CL.20' is the set of numerals between 41 and 60, and so on.

Inspiration for this comes from Honti's (1993) discussion of Finnic overcounting. In Estonian, for instance, 20 is *kaks-kümmend* lit. 'two-ten'. However, the prospective in overcounting is always the ordinal form *teist-kümmend* 'second-ten' (and as in Ingrian, this can shorten to *teist*).

(56) Estonian (Anne Tamm p.c.)
kaks-kümmend
two-ten.PART
'twenty'

(57) Estonian (Anne Tamm p.c.)
kaheksa-teist(-kümmend)
eight-second.PART(-ten.PART)
'eighteen' (lit. eight second (ten))

This pattern is quite general in Finnic. Honti (1993) points out that since the cardinal 20 is *kaks-kümmend* rather than *teist-kümmend*, these overcounting numerals do not make reference to the cardinal 20 itself. Instead, 'the second ten' refers to the entire set of numerals from 11 to 20 ('the first ten' being 1 through 10.) The smaller number is related to this set, and not directly to 20.

---

[20] There is a further context in the Ch'ol NP where A3 morphology occurs, namely ordinals.

(54) Ch'ol (Bale et al. 2019: fn. 20)
Tsa' chäm-i i-cha'-kojty-lel wakax.
PFV die-IV A3-two-CL-NML cow
'The second cow died.'

This is unlikely to be an accident. Ch'ol ordinals are plausibly hidden possessive structures, but I won't attempt to flesh out this proposal here. Instead, I want to raise the question whether the A3 morphology appears in overcounting numerals because the prospective is in an ordinal form. We saw in section 2.3 that there are overcounting languages in which this is the case indeed. With an ordinal prospective, the numeral 24, for instance, would be literally 'four, second score'. This approach might very well work for Classical Kaqchikel (25): in this language ordinals are only marked by the A3 prefix (Brown et al. 2006: 163). In Ch'ol, however, ordinals also bear the *-lel* suffix (Vázquez Álvarez 2011: 159), see (54). As *-lel* is not present in overcounting numerals, the ordinal approach is not supported.

I adopt this insight for Ch'ol, and suggest that the part of (55) containing the prospective is lit. '[the] two-twenty [group] of years'.[21] The whole quantified phrase then corresponds to 'four years in/of/toward [the] two-twenty [group] of years'. The way the two parts of the complex numeral are connected will be made more precise in section 3.3.

That in Mayan overcounting the number sequence is divided into groups of 20 is confirmed by how speakers explain these numerals. In her ethnographic study of numerals in the related Tseltal language (with overcounting numerals built as in Ch'ol), Micalco Mendéz (2010) discusses how the number 20, representing a man with ten fingers and ten toes, is used in counting.

> Naming amounts follows a corporal logic of complete men (groups of 20) implied by a given amount […] Amounts are named by saying the units that remain after having identified the number of complete men. Thus we say that numbers are based on the last "fingers and toes" or units that remain after considering complete men. The remaining fingers and toes are mentioned and then the name of the man to which they belong – the man following the previous complete man. (Micalco Mendéz 2010: 71)

According to Micalco Mendéz's Tseltal consultant, e.g., 199 is lit. '19 10-score' because "It would be 19 units of the tenth man […] It's almost, almost ten men. One unit is missing to get to ten men" (Micalco Mendéz 2010: 71). Thus each interval number is part of a group comprising 20 consecutive numbers.

3.3. THE NATURE OF THE CONNECTION. For additives without an overt morphological connector, e.g., English *twenty two*, it is standard to assume that their syntax is analogous to that of additives with an overt coordinator, e.g., *two hundred and two*. That is, both are &Ps, and the difference lies in whether we have syndetic or asyndetic coordination (Ionin & Matushansky 2018).

I suggest that overcounting languages without an overt morphological connector in general, including Ch'ol, build their complex numerals with the help of a covert P. This P is syntactically and semantically active and has the semantics of 'to(ward)'. It differs from the lative connectors seen in section 2.1 only in not having overt exponence. As shown in section 2, the overt morphological connectors which have been definitively identified in overcounting numerals are of two kinds: Ps and negated verbs. Of these, Ps can be found in various areas of the globe, while negated verbs are restricted to Nagaland, India. Thus in overcounting numerals lacking an overt connector (section 2.3), a covert P is more plausible than a covert V. Putting this together with the previous section's conclusion, I propose (58) as the simplified structure for (48)/(55); xNP stands for 'extended Noun Phrase'.

(58) [$_{xNP}$ chäm-p'ejl ~~jab'~~] P=TO(WARD) [$_{xNP}$ i-cha'-k'al jab']
four-CL year A3-two-CL.20 year
'twenty-four years' (four years in/of/toward [the] two-twenty [group] of years)

The covertness of the P can be viewed as a case of P-drop (as in the Toyota slogan *let's go places*), well-known from varieties of English, German, Italian, Arabic, as well as from Greek and creoles. Across these languages, the omitted P is almost always 'to', and the usual interpretation of the NP whose P has been omitted is a directional goal (Bailey 2018). This meshes well with the fact that in overcounting, the prospective acts as a goal in the direction of which the

[21] Bale et al. (2019: fn. 20) make a similar suggestion in passing: they mention that the prospective may belong to an abstract set of numbers. Their concerns lie elsewhere, however, and they do not elaborate on this point.

counting goes (section 1). Characteristically, P-drop is allowed with (frequently used, familiar or stereotypical) names of places or locations. I suggest that this fits well with assuming a null P in overcounting as well: on the sequence of numbers, multiples of the base correspond to distinguished locations because they function as reference points for computing the location of all other numerals (via addition, subtraction or overcounting). This does not mean that languages in which overcounting numerals lack an overt connector will necessarily have P-drop with motion verbs as well. Complex numerals often allow or require functional elements to remain silent even if this is not possible elsewhere in the grammar, cf. English *twenty two* versus *Mary *(and) Sue*.

Although my main motivation for positing P-drop in Ch'ol (and other languages in section 2.3) is the cross-linguistic frequency with which Ps occur in overcounting numerals, there are two additional considerations why a covert P approach is promising for Ch'ol in particular. Firstly, overcounting in the related Colonial Yucatec Maya language involves an overt P (15), so there is evidence for the presence of this type of structure in the Mayan family. It is attractive to assume that Colonial Yucatec Maya and Ch'ol overcounting numerals are constructed similarly, and the differences can be put down to the exponence of the structure. As in Colonial Yucatec Maya, the covert P of Ch'ol would be *tyi* ('in, on, to, from, by, with'), the all-purpose P of the language. Additionally, as mentioned in footnote 7, some exceptional overcounting numerals in Beltran's Colonial Yucatec Maya data lack the connecting preposition, and are thus identical in their make-up to overcounting numerals in Ch'ol. This minimal variation within Colonial Yucatec Maya is itself best captured as a form of P-drop, I suggest.

Secondly, some multiplicative and additive complex numerals in Ch'ol contain an overt *tyi*. This presents independent evidence that Ch'ol has the wherewithal to build complex quantificational expressions with Ps. We have seen that simplex numerals are immediately followed by a classifier, and that *-k'al* 'twenty' is a classifier itself (46), blocking other, 'regular' classifiers (47). This means that with 20 (or with numerals ending in 20), it is not possible to get classificatory information regarding the animacy, shape, disposition, etc. of the noun from the classifier: *-k'al* is neutral as to these distinctions. In these cases, for the sake of clarity, (Tumbalá) Ch'ol speakers could add *tyi* and the appropriate regular classifier after *-k'al* at least until the 1950s (but this construction does not appear to be productive synchronically). We can understand (59) as 'twenty in animal units'.

(59) Tumbalá Ch'ol (Aulie 1957: 282)
juñ-k'al tyi kojty
one-CL.20 PREP CL
'twenty (animals)'

Although Aulie does not discuss if (59) is restricted to elliptical NPs or it may also occur next to overt nouns (as in Tsotsil), the main point is that *tyi* is attested internally to a complex quantificational expression.

Tyi has also been documented in non-standard numerals. As described in Hopkins et al. (2011), the German ethnographer and historian Karl Sapper collected a Ch'ol word list shortly before the end of the 19th century. One of his Tumbalá Ch'ol consultants occasionally used what Hopkins et al. (2011: 7) characterize as "peculiar numerical expressions" instead of overcounting. Compare the overcounting version of 22 in (60) with the Tumbalá informant's version in (61).

(60) Ch'ol (Bale et al. 2019)
cha'-p'ej i-cha'-k'al
two-CL A3-two-CL.20
'twenty-two' (lit. two, two scores)

(61) Tumbalá Ch'ol (cited in Hopkins et al. 2011: 7, 166)
cha'-p'ejl i-ñumel tyi juñ-k'al
two-CL A3-passing PREP one-CL.20
'twenty-two' (paraphrase in Hopkins et al. 2011: 'two passing the first twenty')

(61) is an additive numeral: it constructs 22 with reference to 20 (*jun-k'al*) rather than 40 (*cha'-k'al*). The connector connected by the nominalized form of the intransitive verb *ñum* 'to pass (by)'. Importantly, the addend is preceded by *tyi* (here probably with a 'from' or 'by' interpretation). While neither (59) nor (61) is an overcounting numeral, these examples show that constructing complex numerals with the help of a P is possible in Ch'ol grammar.

Additionally, in Ch'ol many locative relations and oblique grammatical roles are expressed with relational nouns (Vázquez Álvarez 2011: Ch. 5.7.2). These correspond to axial parts in Svenonius's (2006) terms, and thus can be reasonably assumed to project a PP. Importantly for us, the relational noun *-ik'oty* 'with' is used as a connector of additive numerals in two cases. First, in the 11–19 range addition is the only available strategy. These numerals just juxtapose the augend and the addend, but 16–19 can also optionally feature *-ik'oty* (62). Second, the interval numerals starting from 21 can be expressed not only by overcounting, but also by addition (Aulie 1957; Merrifield 1966). These alternative additives are always constructed with *-ik'oty* (63).

(62) Tumbalá Ch'ol (Aulie 1957: 282)
lujum-p'ejl y-ik'oty wäk-p'ejl
ten-CL A3-RN.with six-CL
'sixteen'

(63) Tumbalá Ch'ol (Aulie 1957: 282)
juñ-k'al y-ik'oty jum-p'ejl
one-CL.20 A3-RN.with one-CL
'twenty-one'

If relational nouns are indeed enveloped in PPs, then additives with *-ik'oty* also support the idea that Ch'ol can build complex numerals with Ps.

Let us turn to how the two NumPs are configured with respect to the P connector in (58). PPs express a relationship between the Figure, the object "whose path or site is conceived as a variable the particular value of which is the salient issue", and the Ground, "a reference-point, having a stationary setting within a reference-frame, with respect to which the FIGURE's path or site receives characterization" (Talmy 1978: 419). In our complex numerals, the smaller numeral corresponds to the Figure and the prospective is the Ground. There is agreement in the literature that the P and the Ground form a constituent to the exclusion of the Figure. This means that the dropped P and the prospective will form a constituent.

(64) [$_{xNP}$ chäm-p'ejl ~~jab'~~] [P=TO(WARD) [$_{xNP}$ i-cha'-k'al jab']]
four-CL year A3-two-CL.20 year
'twenty-four years' (four years in/of/toward [the] two-twenty [group] of years)

What remains to be determined now is how the xNP containing the smaller numeral is related to the rest of the complex numeral. Ionin & Matushansky (2018) argue that there are two strategies for building complex numerals with a P connector. The first one involves adjunction: a PP

containing the P and the Ground is adjoined to the xNP of the Figure. (Again, language-specific rules determine which instance(s) of the lexical noun, if any, will be elided at PF.)

(65) [$_{xNP}$ [$_{xNP}$ NUM N [$_{PP}$ P [$_{xNP}$ NUM N]]]

This is suggested to be the structure of i) additives featuring an 'on' connector (66) or a 'with' connector (not illustrated here), and ii) subtractives with a caritive ('without') connector (67).

(66) Biblical Welsh (Hurford 1975: 603)

un	ar	ddeg
one	on	ten
'eleven		

(67) Welsh Romani (Sampson 1926: 155)

deś	bī	yek'
ten	without	one
'nine'		

Subtractives with an ablative ('from') connector, such as Latin (1), on the other hand, are argued to correspond to PPs, where the smaller numeral's xNP is a degree modifier of the minuend.

(68) [$_{PP}$ [*duo* ~~*pueri*~~] [P=*de* [*sexāgintā pueri*]]]
'two (boys) from sixty boys', i.e., 'fifty-eight boys'

Overcounting is not discussed in Ionin & Matushansky (2018). It would be reasonable to assume that the structure of overcounting numerals featuring a 'to' P is analogous to the structure of subtractives linked by a 'from' P (68). Both involve a connector in the form of a Path adposition and employ an anticipatory pattern (they make reference to the next-higher multiple of the base). The only conspicuous difference between the two types of numerals appears to be the type of Path: subtractives involve a Source Path while overcounting makes use of a Goal Path.

Adapting (68) to Ch'ol overcounting numerals would just require swapping the ablative P to a lative one. This analysis, however, is not tenable for Ch'ol: nouns modified by an overcounting numeral, such as (48) or (49), must have the external distribution of an extended NP rather than a PP. This is because these phrases can act as core arguments. For instance, (49) is (part of) the subject in its clause. Core arguments in Ch'ol are restricted to extended NPs; PPs only introduce non-core arguments (Vázquez Álvarez 2011: 270; Coon 2013: 103). With (68) unavailable, I will adopt the adjunction structure in (65) for Ch'ol (64).

4. Conclusion. This paper offered a cross-linguistic overview of the morphological connectors in overcounting numerals and investigated this type of complex numeral in Ch'ol (Mayan) in more detail. I suggested that Ch'ol overcounting numerals involve two full xNPs (each with its own numeral, classifier and noun) connected by a covert lative adposition.

References

Ambrazas, Vytautas (ed.). 1997. *Lithuanian grammar*. Vilnius: Baltu Lankos.

Andrade, Manuel J. 1955. *A grammar of modern Yucatec.* Microfilm collection of manuscripts on Middle America cultural anthropology series 7, no. 41. Chicago, IL: University of Chicago Library.

Aulie, Wilbur. 1957. High-layered numerals in Chol (Mayan). *International Journal of American Linguistics* 23(4). 281–283.

Avery, John. 1886. The Ao Naga language of Southern Assam. *The American Journal of Philology* 7(3). 344–366.

Bailey, Laura R. 2018. Some characteristics of Southeast English preposition dropping. *Iberia* 10. 48–70.

Bakró-Nagy, Marianne, Katalin Sipőcz & Elena Skribnik. 2022. North Mansi. In Marianne Bakró-Nagy, Johanna Laakso & Elena Skribnik (eds.), *The Oxford guide to the Uralic languages*, 537–564. Oxford: Oxford University Press. https://doi.org/10.1093/OSO/9780198767664.003.0029.
Bale, Alan, Jessica Coon & Nicholás Arcos López. 2019. Classifiers, partitions, and measurements: Exploring the syntax and semantics of sortal classifiers. *Glossa* 4(1). Article 77. https://doi.org/10.5334/gjgl.752.
Beltran, Pedro. 1746 [1859]. *Arte del idioma maya reducido a sucintas reglas*. 2nd edn., José D. Espinosa (ed.). Mérida, Mexico: Imprenta de J. D. Espinosa.
Brown, R. McKenna, Judith M. Maxwell & Walter E. Little. 2006. *La ütz awäch? Introduction to Kaqchikel Maya language*. Austin, TX: University of Texas Press.
Catalogue of Endangered Languages. 2024. Pochuri Naga. University of Hawaii at Manoa. https://www.endangeredlanguages.com/lang/5663.
Chukhapa, K. 2021. The Pochury Naga tribe: A history of their migration and tribe formation. *International Journal of Creative Research Thoughts* 9(5). j761–j775.
Clark, Larry. 1996. The early Turkic and Sarig Yugur counting systems. In Ronald E. Emmerick, Werner Sundermann, Annemarie von Gabain & Sergej Efimovič Malov (eds.), *Turfan, Khotan und Dunhuang,* 17–49. Berlin: Akademie-Verlag.
Clark, Mrs. E. W. (Mary M.). 1893. *Ao Naga grammar with illustrative phrases and vocabulary*. Shillong, India: Assam Secretariat Printing Office.
Comrie, Bernard. 1997. Some problems in the theory and typology of numeral systems. In Palek Bohumil (ed.), *Proceedings of LP'96*, 41–56. Prague: Karolinum – Charles University Press.
Coon, Jessica. 2009. Interrogative possessors and the problem with pied-piping in Chol. *Linguistic Inquiry* 40(1). 165–175. https://doi.org/10.1162/ling.2009.40.1.165.
Coon, Jessica. 2013. *Aspects of split ergativity*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199858743.001.0001.
Coon, Jessica. 2017. Ch'ol. In Judith Aissen, Nora C. England & Roberto Zavala Maldonado (eds.), *The Mayan languages*, 648–684. London/New York: Routledge.
Coupe, Alexander R. 2007. *A grammar of Mongsen Ao.* Berlin: Mouton de Gruyter.
Coupe, Alexander R. 2012. Overcounting numeral systems and their relevance to sub-grouping in the Tibeto-Burman languages of Nagaland. *Language and Linguistics* 13(1). 193–220.
Däbritz, Chris Lasse & Valentin Gusev. 2021. INEL Evenki corpus. Version 1.0. Publication date 2021-12-31. Archived at Universität Hamburg. https://hdl.handle.net/11022/0000-0007-F43C-3. In The INEL corpora of indigenous Northern Eurasian languages. https://hdl.handle.net/11022/0000-0007-F45A-1.
de León Pasquel, María de Lourdes. 1988. *Noun and numeral classifiers in Mixtec and Tzotzil: A referential view*. Sussex: University of Sussex dissertation.
Dékány, Éva. 2022. The syntax of numeral modification in Ainu. Manuscript. HRCL.
Dékány, Éva. 2024. Constituency in classifier expressions: Ch'ol and beyond. *Glossa* 9(1). https://doi.org/10.16995/glossa.10874.
van Driem, George. 2007. South Asia and the Middle East. In Christopher Moseley (ed.), *Encyclopedia of the world's endangered languages*, 283–348. London: Routledge.
Ẹkundayọ, Samuel A. 1977. Vigesimal numeral derivational morphology: Yorùbá grammatical competence epitomized. *Anthropological Linguistics* 19(9). 436–453.
Ehlers, Gerhard. 1983. Notabilia zur alttürkischen Oberstufenzählung. *Ural-Altaische Jarhbücher* 3. 81–87.

Erdal, Marcel. 2004. *A grammar of Old Turkic*. Leiden: Brill.
Fleck, Margaret M. 1981. Tzotzil numeral root morphology. *Journal of Mayan Linguistics* 3. 5–24.
Garczynñski, Błażej. 2014. A short research in Danish cardinal and ordinal numerals on Indo-European background. *Folia Scandinavica Posnaniensia* 16(1). 17–40. https://doi.org/10.1515/fsp-2015-0002.
Greenberg, Joseph H. 1978. Generalizations about numeral systems. In Joseph H. Greenberg (ed.), *Universals of human language*, vol. 3: *Word structure*, 249–295. Stanford, CA: Stanford University Press.
Greenberg, Joseph. 2000. Numeral. In Geert Booij, Christian Lehmann, Joachim Mugdan, Stavros Skopeteas & Wolfgang Kesselheim (eds.), *Morphologie/morphology. Ein Internationales Handbuch zur Flexion und Wortbildung. An international handbook on inflextion and word-formation*, vol. 1, 770–783. Berlin: Walter de Gruyter.
Grünthal, Riho. 2022. Diachronic bottlenecks of the Uralic (ablative-)partitive. *Linguistic Variation* 23(1). 124–156. https://doi.org/10.1075/lv.21003.gru.
Hanke, Thomas. 2005. Overcounting in numerals. LINGUIST List 16.2448. https://v2.linguistlist.org/issues/16/16-2448.html.
He, Chuansheng. 2015. Complex numerals in Mandarin Chinese are constituents. *Lingua* 164 A. 189–240. https://doi.org/10.1016/j.lingua.2015.06.014.
He, Chuansheng. 2022. On the discovery and interpretation of overcounting in Orkhon Inscriptions. *Journal of the Royal Asiatic Society* 33(2). 461–475. https://doi.org/10.1017/s1356186322000256.
He, Chuansheng, One-Soon Her, Xiaoshi Hu & Weijing Zhu. 2017. Overt coordination in additive numerals of minority languages in South China. *Syntax* 20(3). 292–316. https://doi.org/10.1111/synt.12137.
Honti, László. 1993. *Die Grundzahlwörter der uralischen Sprachen*. Budapest: Akadémiai Kiadó.
Hopkins, Nicholas A., J. Kathryn Josserand & Ausencio Cruz Guzmán. 2011. *A historical dictionary of Chol (Mayan); the lexical sources from 1789 to 1935*. Crystal River, FL: Foundation for the Advancement of Mesoamerican Studies. www.famsi.org/mayawriting.dictionary/hopkins/dictionary/Chol.html.
Hurford, James R. 1975. *The linguistic theory of numerals.* Cambridge: Cambridge University Press.
Hurford, James R. 2003. The interaction between numerals and nouns. In Frans Plank(ed.), *Noun Phrase structure in the languages of Europe*, 561–620. Berlin and New York: Mouton.
Hutton, James Henry. 1916. *Rudimentary grammar of the Sema Naga language with vocabulary*. Shillong: Assam Secretariat Print Office.
Ionin, Tania & Ora Matushansky. 2018. *Cardinals: The syntax and semantics of cardinal-containing expressions*. Cambridge, MA: MIT Press.
Jauncey, Dorothy. 2011. Tamabo. In John Lynch, Malcom Ross & Terry Crowley (eds.), *The Oceanic languages,* 608–625. New York: Routledge.
Kaufman, Terrence. 1971. *Tzeltal phonology and morphology*. Berkeley/Los Angeles: University of California Publications.
Lipták, Anikó & Andés Saab. 2016. Movement and deletion after syntax: Licensing by inflection reconsidered. *Studia Linguistica* 70(1). 66–108. https://doi.org/10.1111/stul.12039.

Little, Carol Rose, Mary Moroney & Justin Royer. 2022. Classifiers can be for numerals *or* nouns: Two strategies for numeral modification. *Glossa* 7(1). 1–35. https://doi.org/10.16995/glossa.8437.
Markus, Elena & Fedor Rozhanskiy. 2022a. Ingrian. In Marianne Bakró-Nagy, Johanna Laakso & Elena Skribnik (eds.), *The Oxford guide to the Uralic languages*, 308–329. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198767664.003.0018.
Markus, Elena & Fedor Rozhanskiy. 2022b. Votic. In Marianne Bakró-Nagy, Johanna Laakso & Elena Skribnik (eds.), *The Oxford guide to the Uralic languages*, 330–349. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198767664.003.0019.
Matisoff, James A. 1997. *Sino-Tibetan numeral systems: Prefixes, protoforms and problems.* Canberra: Research School of Pacific and Asian Studies, Australian National University.
Mazaudon, Martine. 1985. Dzongkha number systems. In Suriya Ratanakul, David D. Thomas & Suwilai Premsrirat (eds.), *Southeast Asian linguistic studies presented to André-G. Haudricourt*, 124–157. Bangkok: Institute of Language and Culture for Rural Development, Mahidol University.
Mazaudon, Martine. 2008. Number building in Tibeto-Burman languages. In Stephen Morey & Mark Post (eds.), *North East Indian linguistics*, vol. 2, 117–148. New Delhi: Cambridge University Press India.
McCabe, Robert Blair. 1887. *Outline grammar of the Angami Naga language*. Calcutta: Superintendent of Government Printing.
Menninger, Karl. 1969. *Number words and number symbols: A cultural history of numbers*. Cambridge, MA: MIT Press. English translation of the revised German edition *Zahlword und Ziffer: Eine Kulturgeschichte der Zahlen*.
Merrifield, William R. 1966. Number names in four languages of Mexico. In H. Brandt Corstius (ed.), *Grammars for number names.* 91–102. Dordrecht: Springer. https://doi.org/10.1007/978-94-010-3449-4_6.
Micalco Mendéz, Miriam Moramay. 2010. The vigesimal numeral system: Another logic for constructing and using amounts in the Maya culture's community practices. *Cultural-Historical Psychology* 6(4). 69–74.
Mills, James Phillip. 1926. *The Ao Nagas*. London: Macmillan and Co.
Mills, James Phillip. 1937. *The Rengma Nagas*. London: Macmillan and Co.
Nedjalkov, Igor. 1997. *Evenki.* London: Routledge.
Ochiai, Izumi. 2014. Overcounting in North-Central Vanuatu languages and its distribution. *Kyoto University Linguistic Research* 33. 229–252. https://doi.org/10.14989/196272.
Ochiai, Izumi. 2021. Ainu-go sūsi saikō: Nizyusin-hō-niokeru kahou-sanpou-kar zyōhō-sanpō-e [Ainu numerals revisited: The shift from undercounting to overcounting in the vigesimal system]. *Hoppō Gengo Kenkyū [Northern Language Studies]* 11. 99–121.
Polian, Gilles. 2013. *Gramática del tseltal de Oxchuc*. Mexico City: Centro de Investigaciones y Estudios Superiores en Antropología Social. Publicaciones de la Casa Chata.
Pritsak, Omeljan. 1955. Die Oberstufenzählung im Tungusischen und Jakutischen. *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 105. 184–191.
Sampson, John. 1926. *The dialect of the Gypsies of Wales: Being the older form of British Romani preserved in the speech of the clan of Abram Wood John*. Oxford: Clarendon Press.
Seržant, Ilja A. 2014. The independent partitive genitive in Lithuanian. In Axel Holvoet & Nicole Nau (eds.), *Grammatical relations and their non-canonical encoding in Baltic,* 257–299. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/vargreb.1.07ser.

Simonenko, Alexandra. 2014. Microvariation in Finno-Ugric possessive markers. In Hsin-Lun Huang, Ethan Poole & Amanda Rysling (eds.), *Proceedings of the 43rd Annual Meeting of the North East Linguistic Society*, vol. 2, 127–140. Amherst, MA: GLSA.
Svenonius, Peter. 2006. The emergence of Axial Parts. *Nordlyd: Tromsø University Working Papers in Linguistics* 33(1). 49–77. https://doi.org/10.7557/12.85.
Talmy, Leonard. 1978. Figure and Ground in complex sentences. In Joseph H. Greenberg (ed.), *Universals of human language*, 625–649. Stanford, CA: Stanford University Press. https://doi.org/10.3765/bls.v1i0.2322.
Tamura, Suzuko. 1988/2000. *The Ainu language* ICHEL Linguistic Studies 2. Tokyo: Sanseidō. Originally published in Japanese under the title "Ainugo" (The Ainu language) in Takashi Kamei, Rokurō Kōno, and Chino Eiichi (eds.), *Gengogaku daijiten (Encyclopaedia of linguistics),* 6–94, Tokyo: Sanseidō.
Toland, Norma R. & Donald F. Toland. 1991. *Reference grammar of the Karo/Rawa language*. Ukarumpa, Papua New Guinea: SIL.
Tozzer, Alfred M. 1921. *A Maya grammar, with bibliography and appraisement of the works noted.* Cambridge, MA: Peabody Museum of Archaeology and Ethnology.
Valijärvi, Riitta-Liisa & Lily Kahn. 2017. *North Sámi: An essential grammar*. London: Routledge. https://doi.org/10.4324/9781315733487.
Vasilevič, G. M. 1958. *Èvenkijsko-russkij slovar'*. Moscow: Gosudarstvennoe Izdatel'stvo Inostrannyx i Natsional'nyx Slovarej.
Vázquez Álvarez, Juan Jesús. 2011. *A grammar of Chol, a Mayan language*. Austin, TX: University of Texas at Austin dissertation.
Virtanen, Susanna S., Csilla Horváth & Tamara Merova. 2021. *Pohjoismansin peruskurssin (5 op) [Northern Mansi basic course]*. Helsinki: Helsingin yliopisto.
Žoha, Lukáš, Marcin Wągiel & Pavel Caha. 2022. The morphology of compound numerals: A cross-linguistic case-study. *Linguistics Beyond and Within* 8. 200–217. https://doi.org/10.31743/lingbaw.14966.
Warkentin, Viola & Ruby Scott. 1980. *Gramática Ch'ol*. Mexico City: SIL.
Witter, W. E. 1888. *Outline grammar of the Lhōtā Nāgaā language: With vocabulary and illustrative sentences*. Calcutta: Superintendent of Government Printing.
Wycliffe & La Liga Bíblica. 2010. *Jini nuevo testamento (el nuevo testamento en el ch'ol de Tumbalá) [The New Testament in Tumbalá Ch'ol]*. La Liga Bíblica versión electrónica edn.
Wycliffe & La Liga Bíblica. 2011. *Jini antiguo testamento (la santa bíblia, antiguo testamento en ch'ol de Tumbalá)*. La Liga Bíblica versión electrónica edn.
Yasugi, Yoshiho. 1995. *Native Middle American languages: An areal-typological perspective*. Osaka: National Museum of Ethnology.

Appendix

This appendix provides a synthesis of the languages/language groups mentioned in this study, along with the language family they belong to and the relevant literature on their numerals, organized into groups based on the linker they feature. It is my hope that this will facilitate future work on overcounting languages, regardless of theoretical persuasion.

(69) P connector (adposition or case)
   a. toward: Northern Mansi (Uralic, Honti 1993), varieties of Khanty (Uralic, Honti 1993), non-Southern Saami (Uralic, Honti 1993), Tundra Nenets (Uralic, Honti

1993), Lhōtā Nāgā (Sino-Tibetan, Witter 1888), Nzong (Sino-Tibetan, Mills 1937), Ntenyi (Sino-Tibetan, Mills 1937), Eastern Rengma (Sino-Tibetan, Mills 1937), Southern Hokkaido Ainu (isolate, Ochiai 2021; Dékány 2022), Icelandic (Indo-European, Alexander Pfaff p.c.)

b. genitive: Old Norse (Indo-European, Menninger 1969)

b′. genitive in the half-count: Lithuanian (Indo-European, Ambrazas 1997), Russian (Irina Burukina p.c.)

c. all-purpose P: Classical Yucatec Maya (Mayan, Beltran 1746 [1859])

d. ablative: Evenki (Tungusic, Pritsak 1955), Sakha (Turkic, Pritsak 1955)
[NB: these may be elliptical possessed NPs rather than complex *numerals* per se]

(70) V-based connector: varieties of Ao (Sino-Tibetan, Clark 1893; Avery 1886; Mills 1926; Coupe 2007), Angami (Sino-Tibetan, McCabe 1887), Sema (Sino-Tibetan, Hutton 1916)

(71) No overt connector

a. for (at least some) natural numbers: Old Turkic (Turkic, Ehlers 1983; Clark 1996), Sarig Yugur (Turkic, Clark 1996), Classical K'ich'e (Mayan, Yasugi 1995), Classical Kaqchikel (Mayan, Yasugi 1995), Classical Mam (Mayan, Yasugi 1995), Classical Chontal (Mayan, Yasugi 1995), Tsotsil (Mayan, Yasugi 1995), Tseltal (Mayan, Yasugi 1995), Popti' (Mayan, Yasugi 1995), Chuj (Mayan, Yasugi 1995), Ch'ol (Mayan, Yasugi 1995; Aulie 1957; Bale et al. 2019), Ixil (Mayan, Yasugi 1995), varieties of Zapotec (Oto-Manguean, Yasugi 1995), Central Mansi (Uralic, Honti 1993), varieties of Khanty (Uralic, Honti 1993), Finnic languages (Uralic, Honti 1993), West Tamabo (Austronesian, Jauncey 2011), Northern Amis (Austronesian, Isabelle Bril p.c.), Eastern Rengma (Sino-Tibetan, Mills 1937)

b. only for fractions or complex numerals involving a fractional multiplier (the "half-count"): Classical Danish (Indo-European, Garczyński 2014), German (Indo-European, Garczyński 2014), Faroese (Indo-European, Garczyński 2014), the modern Scandinavian languages (Indo-European, Alexander Pfaff p.c.), Hungarian (Uralic, own knowledge), Lepcha (Sino-Tibetan, Mazaudon 2008)
[NB: Old Norse has the half-count as well as a genitive connector]

(72) To be further investigated

a. for natural numbers in general: Tutuba (Austronesian, Ochiai 2014), Tangoa (Austronesian, Ochiai 2014), Akei (Austronesian, Ochiai 2014), Big Nambas (Austronesian, Ochiai 2014), one of the Ambaen varieties (Austronesian, Ochiai 2014), Paiwan (Austronesian, Ochiai 2014), Central Amis (Austronesian, Ochiai 2014; He et al. 2017), Yami (Austronesian, Ochiai 2014), Ibatan (Austronesian, Ochiai 2014), Ilokano (Austronesian, Ochiai 2014), Tagalog (Austronesian, Ochiai 2014), Kapampangan (Austronesian, Ochiai 2014), Pangasinan (Austronesian, Ochiai 2014), Ibanag (Austronesian, Ochiai 2014), Rawa (Trans-New Guinean, Hanke 2005)

b. for the half-count: Dzongkha (Sino-Tibetan, Mazaudon 1985, 2008), Dungkarpa (Sino-Tibetan, Mazaudon 1985), Bumthang (Sino-Tibetan, Mazaudon 2008), Kaike (Sino-Tibetan, Mazaudon 2008), Gongar (Sino-Tibetan, Mazaudon 2008), Sharchokpa (aka Tshangla, Sino-Tibetan, Mazaudon 2008), Yorùbá (Niger–Congo, Ẹkundayọ 1977), earlier Polish (Indo-European, Garczyński 2014), earlier Russian (Indo-European, Garczyński 2014)

Although in the summary above I aimed to be as objective as possible, a list like this necessarily reflects some analytical choices I made along the way.

I) Contra Hanke (2005), I did not take ordinal morphology to be an "overcounting marker", that is, a linker in the Greenbergian sense (cf. West Tamabo, Finnic languages or Northern Amis, among others, in the main text). This is because in my view, the ordinal marking does not code the syntactic relationship that holds between the prospective and the smaller numeral, it is 'merely' a form/marker of the prospective. I take this to be confirmed by the fact that the ordinal form of the prospective may co-occur with a genuine linker, e.g., a 'toward' P, as in North Saami (8), Icelandic (14) or Kazym Khanty (42), or a genitive linker, as in Old Norse (16). For this reason, overcounting languages which feature the ordinal form of the prospective without a further morphological marker are included in (71).

II) Contra Hanke (2005), I took the "half-count" to be a type of overcounting. Hanke suggests that the "half-count" is just a "related use of division". As pointed out in footnote 2, however, division is multiplication by a fraction, and in languages like Classical Danish, the "half-count" is multiplication by a fraction such that the fraction is itself expressed by overcounting (37). Dzongkha provides evidence that at least in certain languages, the "half-count" does not involve division. In the Dzongkha vigesimal system the "half-count" is used for the odd multiples of ten (73). Numerals which are three quarters on the way to the next multiple of the base (35, 55, etc.) are expressed by what we may call the "three-quarter-count": they involve *ko*, which is a monomorpheme for 'three quarters', the linker *da* and the next-higher multiple of the base (74).[22]

(73) Dzongkha (Mazaudon 2008: 8)
khe pɲhe-da ˈɲiː
score half-da two
'thirty' (~ one and a half scores)

(74) Dzongkha (Mazaudon 2008: 8)
khe ko da ˈɲi
score three.quarter da two
'thirty-five' (~ ¾ into two scores)

(74) cannot involve division. If we divide the second score by three quarters, we will not get 'thirty-five'; the only way to get 'thirty-five' is to multiply the second score by three quarters. As the principle behind (73) and (74) is clearly the same, by parity of reasoning, the "half-count" of Dzongkha does not involve division either.

III) The inclusion of Mayan languages (with the exception of Classical Yucatec Maya) in the "no overt connector" groups reflects my view that the A3 morphology on the prospective does not express a possessive relationship between the two major parts of the complex numeral. Should future work be able to eliminate the problems with the possessive analysis mentioned at the beginning of section 3.2, these languages may be placed under (69b).

[22] Mazaudon (2008) describes *da* as an all-purpose connector, but it is somewhat unclear what this means, as Dzongkha also has case suffixes.