

Chapter 2

Constituency in Cup'ik and the problem of holophrasis

Anthony C. Woodbury

University of Texas at Austin

In Unangan-Yupik-Inuit (UYI) languages, Words are traditionally analyzed as a single Base lexeme, then zero to many Postbases (derivational suffix units), and then inflection according to word class. Since both Bases and Postbases are lexemes that may have concrete meaning, the resulting Word can be phrase-like (holophrastic) even though the languages have no compounding. We evaluate this analysis for verb-headed clauses in Cup'ik, a Central Alaskan Yupik variety, by examining and measuring constituency in the program of the present volume. This yields significant grammatical and phonological confirmation of the traditional Word unit; but the program assumes that the Verb Base will be the single, lexically-dense verb core in a clause, thus not gauging holophrasis, the grouping of multiple lexically-dense elements within a single Word or how such elements might project constituency within or beyond the traditional Word. It is argued that a more complete assay of wordhood within this program must gauge lexical and grammatical contributions to the clause element by element, regardless of Base status. In that way, the program would detect and measure holophrasis as a significant typological dimension along which UYI languages would occupy an extreme position.

1 Introduction

POLYSYNTHETIC languages and polysynthetic constructions are defined as prolifically HOLOPHRASTIC: a single word expresses what in more analytic languages would appear as a whole phrase (Duponceau 1819; see also Boas 1911; Fortescue 1994; and Mithun 2009, who cites Lieber 1853 as the source of the term *holophrasis*).

Mattissen (2004, 2006, 2017), surveying languages considered polysynthetic, finds multiple elements within a putative word unit carrying lexical meaning, and encoding such categories as “event or participant classification and quantification, setting (e.g., ‘in the night’), location or direction, motion, instrument (e.g., ‘by hand’)”. The survey points to considerable diversity in the kinds of lexical meanings and categories that can be expressed morphologically and in which ones of these are selected from language to language; and even considerable diversity in how such categories are encoded: as roots alone, roots within compounds, clitics, affixes, featural ablaut/mutation, or as combinations of any of these. Even more fundamentally, holophrasis presupposes a theoretically and empirically stable notion of the *WORD*; and yet the word is what this volume aims to scrutinize.

Unangan-Yupik-Inuit (UYI) languages (historically termed Eskimo-Aleut languages) pose the problem interestingly: constituency diagnostics are quite agreed that Words—where capitalization signals a formulation optimized to account succinctly for a language’s own patterns—are made of a single Base lexeme and zero or more Postbases (derivational suffix lexemes): so every time you encounter a new Base, you start a new Word. And since Postbases are lexemes, often quite concrete and often heading productive constructions that grammatically extend beyond the putative Word in which they occur, they give a clear impression that UYI Words are holophrastic, even though—as it happens—there is no actual Base plus Base compounding.

The purpose of this chapter is to examine and measure constituency in Cup’ik verb-headed simple clauses in terms of a comparative program that is designed for that purpose, namely that of Tallman (2021) and the chapters in the present volume. Cup’ik, spoken in Chevak, Alaska, is a variety of Central Alaskan Yupik (CAY), which in turn may serve as a typical representative of the UYI family, in particular its Yupik-Inuit (YI) branch. I will then evaluate the results by comparing them against the largely agreed-on historic framework that was just mentioned in which Cup’ik and other UYI languages have been analyzed, and on that basis evaluate and critique our comparative program. In particular I will show that the program perspicuously demonstrates the depth and breadth of evidence for the long and complex Word unit posited for the family traditionally; but that as formulated, it comes up short in that it does not fully measure the *LEXICAL DENSITY* of combining elements; that is, roughly, where those elements sit along a continuum from maximally content-bearing and lexical, to non-content-bearing and grammatical-only. Such a continuum is expressed by Sapir (1921: 106–126) in his scalar framing of what he calls *BASIC VS. RELATIONAL CONCEPTS*, and their

(defeasible) tendency to align, respectively, with morphological roots vs. grammatical processes like affixation, ablaut, and so on. Relatedly Croft (2001: 244) presents a notion of PRIMARY INFORMATION BEARING UNIT within a constituent as part of a consideration of headedness that we will return to later.

Rather than measure the lexical density of combining elements, the present program requires that Cup'ik Verb Bases (roots or stems) should serve as the lexically-dense constituent anchor or VERB CORE in Verb Words, rather than Postbases (suffixes or suffix clusters), even if some Postbases should turn out to be functionally verb-like and have considerable lexical density. Because of that, it does not allow consideration for the phenomena that have led to perceptions of holophrasis, namely the grouping of multiple lexically-dense pieces into a single word and the recognition of how such elements might project constituency apart from their participation in traditionally-recognized holophrastic words. I argue, then, that a more complete assay of wordhood within this program must actually gauge constructional contributions—including lexical density—element by element, whether it is a root/stem or not. In that way, it can be possible to detect and measure the presence, degree, and impact of holophrasis.

2 Cup'ik and the Unangan-Yupik-Inuit languages

Cup'ik, as noted, belongs to the Unangan-Yupik-Inuit family, whose genetic subgrouping is as shown in (1), based on Woodbury (1984), Fortescue et al. (2010), and the discussion there:

- (1) Unangan-Yupik-Inuit (UYI) language family:
 - Unangan [Formerly Aleut] (Eastern and Western varieties)
 - Yupik-Inuit [Formerly Eskimo]
 - Sirenik (Russia; presently dormant)
 - Yupik
 - * Siberian Yupik (2 languages)
 - Central Siberian Yupik (Russia and Alaska)
 - Naukan Siberian Yupik (Russia)
 - * Alaskan Yupik (2 languages)
 - Central Alaskan Yupik (CAY): Cup'ik (Chevak); Cup'ig (Nunivak Island); General Central Yupik varieties (Yukon Delta to Bristol Bay, variously known as Yup'ik, Yupiaq, Yugtun); others

- Pacific Yupik (several varieties)
- Inuit-Iñupiaq (Dialect continuum from Northern Alaska to Canada to Greenland; varieties known as Iñupiaq (Alaska); Inuvialuktun, Inuktitut, Inuttut (Canada); Kalaallisut (West Greenland), among others

Major sources on CAY include several grammars (Reed et al. 1977; Jacobson 1995 and Miyaoka 2012) and an outstanding dictionary (Jacobson 2012). Woodbury (1981) focuses on the Cup'ik variety in particular. All of this work continues a broader framework established and expanded in writings on Kalaallisut, especially Kleinschmidt (1851), Bergsland (1955), Fortescue (1984) and Sadock (2003). The tradition is also notably expanded in de de Reuse (1994), which focuses on Central Siberian Yupik. Despite many differences in interpretation, all this work shares a deep commitment to describing the languages on their own terms; and the agreement across these works is one of many indications of a unique and also shared typological build across the languages themselves, as I discuss in Woodbury (2017). The present chapter is based on about 20 hours of transcribed naturalistic speech and a large number of sentences from elicitation, all created over the period from 1978 to 1997 in visits to Chevak, Alaska and mostly archived at the Alaska Native Language Archive in Fairbanks, Alaska.

3 The Cup'ik word in own-terms description

In this section I outline those aspects of Cup'ik relevant for this article in a way that strives to fit the language's own patterns and that also largely conforms to the traditionally-established framework for UYI grammar. I term this OWN-TERMS DESCRIPTION, repeating the oft-repeated adage, but I do not mean to suggest that a single most optimal own-terms account always exists. Rather, I'm suggesting a stance that values internal perspicuity over extrinsic plans or frameworks (see Haspelmath 2011 for some more concrete discussion). In what follows I draw heavily on Woodbury (1981 and 2017), which give more detail. We'll consider first the architecture of (traditional) Words (inflection, derivation, and sometimes clitics); and then briefly review features that support the characterization of UYI Words as holophrastic.

3.1 Inflectional morphology

Based on inflectional and other patterns, Cup'ik has three classes of WORD: NOMINALS, VERBS, and PARTICLES (again, note the use of capitalization because these terms are in some ways parochial).

Nominal Words are a super-category consisting of common nouns, independent pronouns, demonstratives, quantifiers, participles, and others, and they are inflected for case, number, and (for common nouns) the person and number of any possessor. For example, (2) shows a noun phrase consisting of two Nominal Words, a demonstrative and a noun, both inflected for Absolutive singular, where either constituent is optional (marked by parentheses); and (3) shows several possessor-possessum constructions, where the case value of the possessor, always syntactically optional, is the Relative case while the inflection of the possessum reflects the person and number of the possessor in addition to its own number and case.¹

(2) Demonstrative – Noun

(*tau-na*) (*arnaq*)
(that-ABS.SG) (woman.ABS.SG)
'that woman; woman; that one.'

(3) Possessor – Possessum

- a. (*arna-m* *tau-m*) *eni-i*
(woman-REL.SG that-REL.SG) house-ABS.SG+3SG.POSS
'that woman's house; her house.'
- b. (*wii*) *en-ka*
(me.REL.SG) house-ABS.SG+1SG.POSS
'my house (of mine)'

¹Cup'ik has these phonemes: /p,t,tʃ,k,q,m,n,n̄,ŋ,v,f,l,ɬ,j,s,ʃ,ɣ^w,x,x^w,ɬ,ɬ^w,χ,χ^w,a,i,u,ə/ (Woodbury 1981). Cup'ik examples are cited in the standard Central Alaskan Yupik orthography (Jacobson 2012), where all symbols represent phonemes with the same IPA value except: <vv> = /f/; <ll> = /t/; <gg> = /x/; <ww> = /x^w/; <rr> = /χ/; <ūr> = /ɬ^w/; <urr> = /χ^w/; <ng> = /ŋ/; <ín> = /m̄/; <ñ> = /n̄/; <ńg> = /ŋ/; <c> = /tʃ/; <y> = /j/; <g> = /ɣ/; <w> = /ɣ^w/; <r> = /ɬ/; <e> = /ə/; and apostrophe following a consonant symbol and preceding a vowel symbol <C'V> indicates that the consonant is phonemically geminate /C:V/. Also, voiced continuant symbols represent their voiceless counterparts in clusters with other voiceless sounds, hence <maligtellruanga> 's/he followed me' represents strictly-phonemic /malixtəɬxuana/, where orthographic <g> and <r> are representing the voiceless phonemes /x/ and /χ/. Furthermore, when examples are segmented morphologically, the segmentations are performed on the orthographic (and therefore phonemic level) spelling, and for that reason it will become obvious to the reader that quite extensive morphophonological rules are at play, so that a Base or Postbase we are discussing may show up with different shapes in different contexts. Some idea of these rules is outlined in §5.7.2, when we take up the segmental phonological basis for constituency; but the curious reader will find full discussions of these matters in Woodbury (1981) for Cup'ik and of very similar facts for other Central Alaskan Yupik varieties in Miyaoka (2012) and Jacobson (2012). In any case, I have chosen not to add a regularized, morphophonemic line to each example because for present purposes, it does not add much, and it's also rather cumbersome.

- c. (*taluya-m*) *quli-ini*
 (trap-REL.SG) area.above-LOC.SG+3SG.POSS
 ‘above the trap; above it.’ (lit: At the trap’s area-above)

Verb Words likewise are inflected, but for one of ten or so values for Mood, which indicates illocutionary force or type of subordination for the clause of the Verb Word in which it appears; and then for the person and number of the surface intransitive subject (S), transitive object (O), and, in most Moods, also the transitive subject (A). (4a)-(4b) show intransitive clauses consisting of a Nominal Word S in the Absolutive Case, which is always optional when recoverable; and a Verb Word in the Indicative and Appositional Moods; (5a)-(5b) shows transitive sentences with Nominal Word A in the so-called Relative Case² and O in the Absolutive Case, again both optional, and Indicative (5a) and Appositional Mood (5b) verbs agreeing in person and number with O, and in (5a) but not (5b) also with A because the Indicative Mood requires A-agreement while the Appositional Mood excludes it.³

(4) S – V

- a. (*Arnaq*) *qavar-tu-q.*
 woman.ABS.SG sleep-IND-3SG.S
 ‘The woman/She is sleeping.’
 b. (*Wangkuta*) *qavar-lu-ta.*
 we.ABS.PL sleep-APPOS-1PL.S
 ‘We, sleeping.’

(5) A – O – V

- a. (*Arna-m*) (*kaugpii-t*) *tangrr-a-i.*
 woman-REL.SG walrus-ABS.PL see-IND.3SG.A-3PL.OBJ
 ‘The woman/She saw the walruses/them.’
 b. (*Kaugpii-m*) (*wii*) *tangrr-lu-a.*
 walrus-REL.SG me.ABS.SG see-APPOS-1SG.OBJ
 ‘The walrus/It, seeing me.’

²RELATIVE is the traditional term for this case, although some more recent writings on Inuit varieties outside Alaska use the term ERGATIVE. Relative persists in part because it marks possessors in addition to just transitive subjects.

³The APPPOSITIONAL Mood—with mood sign *-lu-* ~ *-na-* in all YI languages and known in the literature also as CONTEMPORATIVE, CONJUNCTIVE, or SUBORDINATIVE—indicates a clause in apposition or co-subordination with another clause, with which it normally shares a S/A subject. It can usually be glossed in English with a present participle, as I have done in (4b) and (5b).

Finally, Particle Words are Words that lack inflection. They function as adverbs, discourse particles, and interjections, e.g., *unuagu* 'tomorrow', *cali* 'more', *ataam* 'again', *wall'u* 'or else,' and *Aren!* 'Oops!'.

As should be clear from the examples, the inflectional categories marked on Nominal Words and on Verb Words are expressed by a suffix or suffix cluster we can, as a whole, term the NOMINAL INFLECTION and VERB INFLECTION. While they systematically encode the category values described above, they often do so in a way that is non-biunique: for example, Absolutive Singular is *-na* for Demonstratives and otherwise nothing, as shown in (2); *-tu-* 'Indicative' in (4a) shares nothing with *-a-* 'Indicative third singular A' in (5a); whereas in (4b)-(5b) *-lu-* is a more consistent marker of the Appositional Mood. Yet nearly always, the inflectional marking is linearly distinct, as suffix or suffix cluster, from its Nominal or Verb host, which we can identify as the NOMINAL BASE and the VERB BASE. This can be expressed as rules in (6), where Inflection designates whatever suffix or suffix bundle expresses the appropriate obligatory inflectional categories, indicated in (7):

(6) Inflection rules

- Nominal Word → Nominal Base + Nominal Inflection
- Verb Word → Verb Base + Verb Inflection

(7) Inflectional categories

- Nominal Inflection
 - Number (Singular, Plural, Dual)
 - Case (Absolutive, Relative, Obliques)
 - Possessor Person (1,2,3,Reflexive) and Number
- Verb Inflection
 - Mood (Indicative, Interrogative, Optative, Appositional two Participial Moods (transitive and intransitive), and five or so Adverbial Moods with values like 'when in the past', 'whenever', 'while', 'if/when hypothetically', and 'although').
 - Person/number of surface S, O, A

We can also make two typological observations at this point. First, insofar as nominal inflection includes information about external possessors and verb inflection includes information about S, O, and A, the pattern is one of HEAD-MARKING. But since possessor, S, O, and A NPs are also marked for case, the pattern is also one of DEPENDENT-MARKING. Thus, Cup'ik displays what Nichols

(1986) has termed **DOUBLE-MARKING**. Second, in terms of alignment, nominal case-marking in nominals is mostly **ERGATIVE-ABSOLUTIVE** and virtually never **NOMINATIVE-ACCUSATIVE**, whereas the complex and often non-biunique patterns within verb inflection show both alignments depending on mood (Woodbury 1981: 141–189).

3.2 Derivational morphology

We now take up the composition of Nominal and Verb Bases (but leave aside that of Particle Words). Nominal and Verb Bases can be simple lexemes, representable as lemmas in a lexicon: Jacobson (2012: 12)’s CAY dictionary (by its own count) lists 11,200 of them.

3.2.1 Base recursion

But more complex Nominal Bases and Verb Bases can be derived by a simple recursive process, spelled out in (8):

- (8) **BASE RECURSION RULE**
Base \rightarrow Base + Postbase

POSTBASE is a term of art first arising in Reed et al. (1977) that refers to a suffixal lexeme which selects either a Nominal or a Verb Base, and from it, derives either a Nominal or a Verb Base. There are therefore four major classes of Postbase: those both selecting and deriving a Nominal Base (NN); those both selecting and deriving a Verb Base (VV); those selecting a Verb Base and deriving from it a Nominal Base (VN); and those selecting a Nominal Base and deriving from it a Verb Base (NV).⁴ Of these Jacobson (2012)’s CAY dictionary lists about 540, some but not all of which are fully productive; and none of which can also function as Bases.⁵ Furthermore, it is possible to keep adding Postbases to an ever growing complex Base, as long as the selectional criteria are observed. Strings of up to five Postbases are not uncommon. Examples (9) and (10), from Words occurring in Cup’ik texts, give the flavor:

⁴There also are a few further minor possibilities, including the selection of a Particle, the selection of an inflected Word (Woodbury 1996, Sadock 2017), and the derivation of a Particle Word.

⁵Even at the level of the UYI family, Bases reconstruct as Bases (or Bases plus suffixes) and Postbases as suffixes or suffix clusters, with only a tiny class of exceptions (Fortescue et al. 2010). Thus there is virtually no reanalysis (or ‘grammaticalization’) of Bases as Postbases or vice versa: the classes are disjunct.

- (9) Woodbury (2017: 542)
- | | |
|---------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------|
| ivruci- _N | 'waterboot (N)' |
| ivruci- li - _V | ' make waterboots (for)' (-li- NV 'make (for)') |
| ivruci-li- ste - _N | ' one who makes waterboot (for)' (-ste- VN '(possessor's) one who does V (tr)') |
| ivruci-li-ste- ngerr - _V | ' have someone who makes (one) waterboots' (-ngqerr- NV 'have') |
| ivruci-li-ste-ngqer- sugnaite - _V | ' definitely not have someone who makes (one) waterboots' (-yugnaite- VV 'definitely not') |
| ivruci-li-ste-ngqer-sugnail- ngur - _N | ' one that definitely doesn't have someone who makes (his/her) waterboots' (-ngur- VN 'one who does V (intr.)') |
- (10) Woodbury (2017: 542)
- | | |
|-------------------------------------------------------|------------------------------------------------------------------------|
| quuyurni- _V | 'be smiling' |
| quuyurni- arte - _V | ' suddenly be smiling' (-arte- 'suddenly') |
| quuyurni-arte- llru - _V | 'suddenly smiled' (-llru- VV 'did') |
| quuyurni-arte-llru- yaaqe - _V | 'suddenly smiled, but alas ' (-yaaqe- VV 'alas') |
| quuyurni-arte-llru-yaaqe- llini - _V | ' evidently suddenly smiled, but alas' (-llini- VV 'evidently') |

(9) demonstrates the possibility of PING-PONG RECATEGORIZATION (Mattissen 2017: 86), deriving back and forth between Nominal and Verb bases; while (10) shows the continuous elaboration of a verb base. As also can be seen, there appears to be a semantic corollary to this recursive process that we can formulate as follows:

(11) POSTBASE SCOPE RULE

A Postbase has scope over exactly the base it selects.

To the extent (11) is true (although it isn't entirely true, as we're about to see), the whole Base to the left of a given Postbase is a constituent, semantically speaking.

3.2.2 Templatic pre-inflection

A wrinkle in the story just told is that a small number of VV Postbases align in a fixed order just before the Verb Inflection, shown in (12) and illustrated in (13), which I have called the TEMPLATIC PRE-INFLECTION (Woodbury 1981), and which follows patterns that might largely be predicted in terms of scope (e.g., Foley & Van Valin 1984: 208–224):

- (12) Order: 1 < 2 < 3 < 4 < 5 < 6 < 7 < Verb Inflection (where the elements in each slot are mutually exclusive)
1. ASPECT: *-tu-* ‘always’, *-yuite-* ‘never’, *-qar-* ‘momentarily’
 2. REALIZATION1: *-yaaqe-* ‘in vain’, *-ngate-* ‘seem’, *-ksaite-* ‘not yet’
 3. TENSE: *-llru-* ‘did’, *-ciqe-* ‘will’, *-ngaite-* ‘won’t’
 4. STATUS: *-nrite-* ‘not’; *-yugnarqe-* ‘may’; *-yugnaite-* ‘definitely won’t’
 5. REALIZATION2: *-yaaqe-* ‘in vain’, *-ngate-* ‘seem’
 6. EVIDENTIAL: *-llini-* ‘evidently’; *-lli-* ‘perhaps’
 7. TENSE-MODAL: *-(g)aqe-* ‘would have’, *-ki-* (requires Optative) ‘will; did (narrative reading)’; *-lqe-* (requires Indicative transitives) ‘did’)
- (13) a. 3-6
melu-llru-llini-u-q
 smoke-did-evidently-IND-3SG.s
 ‘evidently s/he smoked.’
- b. *6-3
 **melu-llini-llru-u-q*
- (14) 3-6-7
pi-llru-lli-aqe-ka-it
 do-did-perhaps-would.have-TRPRT-3PL.A+3PL.OBJ
 ‘that they would have maybe told them.’
- (15) 1-3-6
liica-tu-llru-lli-ki-it
 teach-always-did-perhaps-TRPRT-3PL.A+3SG.OBJ
 ‘that they *maybe used to* teach him.’
- (16) 3-4
pi-vaka-llru-nril-ke-ka
 do-fully-did-not-TRPRT-1SG.A+3SG.OBJ
 ‘that I didn’t fully [obey] it.’

(17) a. 2-3

naptar-c-aaqe-llru-u-nga

whitefish-catch-in.vain-did-IND-1SG.s

'I caught a whitefish, but alas.', e.g., it got away

b. 3-4

naptar-te-llru-yaaq-u-a

whitefish-catch-did-in.vain-IND-1SG.s

'I did alas catch some whitefish.', e.g., a veiled offer

(13) and (17) are elicited forms that test alternative orderings; (14)-(16) are text examples that further illustrate the ordering claimed in (12). Such fixed ordering--albeit with substantial differences in detail--is found before verb inflection in all Yupik-Inuit languages (hence the term *Templatic Pre-Inflection*) and has been treated as a departure from the strictly binary, right-branching model specified by the Base rule (8). For example, some have argued that the ordered VV Postbases form branching constituents (Fortescue 1980 for Kalaallisut) that sometimes also include the Verb Inflection (Woodbury 1981 for Cup'ik, de Reuse 1994 for Central Siberian Yupik). It is also possible instead to leave (8) alone, but to impose the template implied by the formulation in (12) as a filter. It would seem that templatic ordering would weaken the Postbase Scope Rule (11), but not entirely: as can be seen in (12), the Realization Postbases are "wild cards" that may occur in two positions (2 and 4), with typical scopal effects as in (17a)-(17b), see Woodbury (2017: 554-555).

3.3 Enclitics

Certain particles are treated as enclitics, occurring in a mostly fixed order at the ends of Words of any class. They are sometimes treated as regular parts of Words but often marked distinctly--e.g., with clitic boundaries noted orthographically--so as to form what we can call *Clitic Groups*, shown in (18), where 'Enclitics' stands for a sequence of from one to four Enclitics following a specific ordering and never numbering more than four (see Woodbury 1981:292-294), and illustrated in (19), where '=' marks Enclitic boundaries:

(18) CLITIC GROUP RULE

Clitic Group → Word + (Enclitics)

(19) *Tayima=llu=ggur=am* *pii-nani*

elsewhere=and=it.is.said=again absent-APPOS.3.REFL.SG.S

'and it is said, again, he was absent, somewhere else.'

In (19) the host is a Particle Word *tayima* ‘elsewhere’ and three Enclitics appear on the first Word in the whole host phrase, as is frequent. The Clitic Group itself is based in part on the fixity of clitic order but also on its status as a superordinate domain for automatic stress rules, albeit with a few differences from those stress rules applying to the Word proper (see §5.7.1 for further discussion).

3.4 The case for holophrasis

The foregoing lets us articulate two ways that Cup’ik is holophrastic, that is, that a single word expresses what in more analytic languages would appear as a whole phrase. One is the lexemic character of both Bases and Postbases; the other is the propensity of certain VV Postbases to head phrase-like or even clause-like constituents that go beyond just the Base to which they are suffixed. These are taken up in turn in the following two subsections.

3.4.1 The complex lexemic character of bases and Postbases

Both Bases and Postbases are lexemes, that is, elements that are PRODUCTIVE grammatical and semantic formatives in larger constructions, and they are LISTEMES, that is, elements with non-compositional meaning that therefore have to be listed in the lexicon. (However, because Base formation (8) is recursive, we are really speaking here only about Bases not recursively formed as the output of (8)). Such Bases then are the productive listemes consisting of the idiomatic collocation of a root or stem, and zero or more less-than-productive suffix elements. Thus, the Nominal Base *qayaq* ‘kayak’ consists of a single root *qayaq* whereas *ivruciq* ‘water boot’, as in (9), consists of a semi-idiomatic combination of the verb Base *iver-* ‘to wade’ plus a suffix (also a Postbase) *-(u)ciq* ‘means for V-ing’; but the meaning is not just any device for wading, but specifically a thigh-high fish-skin or seal-gut boot. Postbases likewise are productive lexemes, but consisting of one or more suffixal elements. For example, drawing again on (9), *-li-* is a NV Postbase meaning ‘to make V’ that consists only of the suffixal element *-li-*. That same suffixal element can also combine with another suffixal element, *-ur-*, usually having ‘habitual’ meaning, to form a complex Postbase *-liur-* ‘to deal with N’, which productively selects a Nominal base to form a Verb Base. But unlike the *-li-* suffixal element, the *-ur-* ‘habitual’ suffixal element is not a stand-alone productive Postbase; and moreover, the whole meaning of *-liur-* ‘to deal with N’ is somewhat semantically idiomatic and non-compositional. So *-liur-* is a productive suffixal listeme and lexeme, what we’re calling a Postbase. An even more complex, but quite typical Postbase from (9) is *-yugnaite-* ‘to definitely not

(do) V', composed of the elements *-yug-*, which as an independent VV Postbase means 'to want V'; plus *-nar-*, not an independent Postbase but occurring within some Postbases with the meaning 'to tend to V'; and then *-(ng)ite-*, which independently is an NV Postbase meaning 'to lack N'. Although this may compose into something like 'to lack a tendency to want to V', the actual meaning, 'to definitely not (do) V', is hardly the same. In short, etymologically complex Postbases, like etymologically complex Base lexemes, are not the semantic sums of their parts.

The other part of the equation—as seen in the discussion of rule (8)—is that as lexemes, Bases and Postbases do indeed work productively and compositionally. Writing about Central Siberian Yupik, de Reuse (2009) proposes the term **PRODUCTIVE NONINFLECTIONAL CONCATENATION (PNC)**. He argues that it is a hallmark of polysynthesis, and that it is especially elaborate in Yupik-Inuit languages, for which it offers a specific mechanism for the vaguer notion 'holophrasis'.

3.4.2 The lexical density of certain Postbases

A second strand of the argument for holophrasis is establishing the lexical density of the elements within words: this is germane because holophrasis is not holophrasis if the components of Words are not word-like; and part of being word-like is having lexical density. For UYI languages, the lexical density of Nominal and Verb Base lexemes is established and can be assumed. What is worth establishing is the lexical density of Postbases. What follows is a synoptic survey of VV Postbase meanings (a far deeper and more exhaustive account is in Miyaoka 2012 for Yup'ik). It is broken into two groups: Postbases acting as heads that select Verb Base as complement; and Postbases acting as adverb-like modifiers to the Verb Bases they morphologically select (quoted from Woodbury 2017: 545):

VV Postbase-as-head (verb-selecting verb). (i) causative and other transitive, complement-taking, argument-structure affecting verbs (Sec. 30.7.2 further characterizes this class, called double transitives): 'let', 'ask/tell to', 'say that', 'think that', 'wait for'; (ii) other argument structure-affecting verbs, auxiliaries, or voice markers: 'to do V-tr. (to)' (antipassive, suppresses O or makes it oblique), 'to do V in place of', 'to do V on account of' (applicative), 'for S (pl.) to do V to each other', 'tend to cause V (intr.)', 'be time (for O or S) to V or be V-ed', 'will/should V or be V-ed', 'be more V (stative) than (oblique)', 'test how V (stative) O is'; (iii) verbs of ability: 'be able to', 'be ready to', 'not any longer be able to'; 'be able to do V proficiently'; (iv) verbs of desire, propensity, purpose, or modality: 'want to', 'want O to',

‘tend to’, ‘no longer care to’, ‘be ready to’, ‘be ready at any moment to’, ‘in order to’, ‘be about to’, ‘plan to’, ‘must/should’ (v) verbs of endeavor: ‘try to’, ‘try unsuccessfully to’ (vi) verbs indicating phases of accomplishment: ‘begin to’, ‘be about to’, ‘set out to’, ‘go and V’, ‘be in a state of V-ing’, ‘to become V (stative)’, ‘to reach a state of V (stative)’; ‘stop V-ing’.

VV Postbase-as-modifier (adverb). (i) manner adverbs: ‘poorly’, ‘happily’, ‘well’, ‘easily’, ‘roughly’, ‘quickly’; (ii) adverbs of degree, speed, and intensity: ‘more and more’, ‘excessively’, ‘intensely’, ‘really’, ‘suddenly’, ‘barely’, ‘a lot’, ‘a little’, ‘just, only’, ‘almost’; (iii) affective epithets: (Sec. 30.7.3 further characterizes this class, which usually modify verbal actants): ‘poor’, ‘darned’, ‘young’, ‘dear’; (iv) relators to other events: ‘first’, ‘also’, ‘again’, ‘never again’, ‘finally’, ‘earlier’, ‘later’; (v) aspect-related adverbs: ‘do V to O (pl) one after another in succession’, ‘continuously’, ‘now and again’, ‘habitually’, ‘customarily’, ‘always’, ‘first’, ‘repeatedly V and un-V’; (vi) negators: ‘not’, ‘will not’, ‘not yet’, ‘don’t V!’ (vii) tense markers: ‘in the past’, ‘in the future’, ‘not in the future’; (viii) markers of propositional attitude: ‘evidently’, ‘contrary to expectation’, ‘authentically’, ‘probably’, ‘but alas’, ‘maybe’, ‘seemingly, perhaps’ ‘probably’, ‘definitely not’; (Sec. 3.2.2 discusses special ordering properties of some of v-viii).

Not all VV Postbase meanings described here are equally lexically dense, but even the sheer number of meanings suggests that at least some have very considerable lexical density. Correspondingly, the sheer number of VV Postbases makes them as a whole an open rather than a closed class (even if certain Postbases are more grammatical than lexical). It is also worth noticing the complete lack of adverbial meanings based on body part as instrument (‘by hand’), location (‘above’), setting (‘on the beach’), direction (‘toward speaker’), specific time (‘at night’) that are otherwise common in polysynthetic languages, cf. Mattissen 2017.

3.4.3 The syntactic independence of certain Postbases

In his article, de Reuse also argues that PNC—unlike derivational morphology—‘interacts with syntax’ in a way that would violate the Postbase Scope Rule (11). By ‘interacts with syntax,’ he refers to a long set of debates, the crux of which is the extent to which certain productive verb-deriving (i.e., NV and VV) Postbases—can be treated as ‘syntactic atoms’ at some level, even as they function in the morphological treatment just outlined as PNC’s (Sadock 1980, 1991, Woodbury

& Sadock 1986, Mithun 1984, Baker 1988, de Reuse 1994, Johns 2007, Compton & Pittman 2010, Fortescue 2015, Yuan 2018). We will take up this issue in more detail in §6, but consider the following example:

(20) (Woodbury 2017: 352)

[*ciku-meng atauci-meng ene*]-*ngqer-tu-a*

ice-INS.SG one-INS.SG house-have-IND-3SG.S

'I have one house made of ice.'

Here the NV Postbase *-ngqerr-* 'to have N' seems to take as its complement not just its Nominal Base host *ene-* 'house', but (at some, possibly abstract level, represented with square brackets) an NP meaning 'one house made of ice' that is expressed, in part, by the stranded modifiers *cikumeng* '(with) ice' and *atauci-meng* '(with) one'. The question then is the extent to which *-ngqerr-* 'to have N' functions as a verb, and the extent or level at which 'one house made of ice' is a constituent. Although a lot of the debate is expressed in the terms of popular cross-linguistic frameworks (Haspelmath 2011), there is widespread agreement on the 'own terms' basics of constructions such as these.

3.4.4 Conclusion

The crux of UYI specialists' own-terms analysis, even across internal interpretive divides, is that these languages have highly complex Words built, productively, from lexemic Bases; (lexemic) Postbases; inflection; and phonologically adjoined Clitics. At the same time, they recognize that these Word components bear resemblances to units treated as words in other languages insofar as the components can themselves behave as 'syntactic atoms' and can often have morphologically complex and idiomatic internal composition.

With all of that in mind, our task now is to measure UYI languages—taking Cup'ik as the case in point—in cross-linguistic comparative terms, first, for the obvious end of comparing it in a consistent way to what is becoming an impressive sample of other languages measured similarly; but second, in order to see to what extent the comparative program captures all of what has arisen in the centuries-old UYI linguistic tradition, congratulating it where it does so, but proposing revisions or expansions or clarifications where it does not do so.

4 A planar structure to diagnose constituency in the Cup'ik clause

The PLANAR STRUCTURE FOR THE VERB is provided in Table 1. It is based on flattening out and elaborating the generalizations from §3. In the system we are using, every planar structure is lexically anchored, so to speak, on a lexemic CORE element that is obligatorily present. For Tallman, that core or anchor must be a root or stem, relative to which the surrounding planar positions are located. Accordingly he defines the VERB CORE “as a verb root *or* as a verb stem which would no longer remain of the same category if any of its affixes were stripped of (Tallman 2021: 13). Applying the first part of this criterion to Cup'ik, then, we can consider Verb Base from §3 (before the recursive application of the Postbase rule in (8)) as the verb core in terms of which the rest of the verbal planar structure is defined. Since—as noted—lexemic Verb Bases are not necessarily unanalyzable roots, we will engage in a certain amount of sleight of hand by considering lexemic Verb Bases as the verb core in position 2 even if they aren't single simple roots, but rather consist of a root or stem, together with less-than-productive etymological VV suffixes, e.g., *eliynga-* ‘to be knowledgeable’ from *elite-* ‘to learn’ plus the restricted and only semi-productive VV suffix *-nga-* ‘be in a state of having ‘V-ed’.⁶ Applying the second part of Tallman's criterion, ‘a verb stem which would no longer remain of the same category if any of its affixes were stripped off’, is more straightforward. For example, if a derived Verb Base such as *ene-ngqerr-* ‘house-to.have.N’ = ‘to have a house’ were formed from a Noun Base (*ene-* ‘house’) and a NV Postbase (*-ngqerr-* ‘to have N’), the whole derived Verb Base would count as a Verb Core and thus would fill just position 2 in our planar structure.

From there, VV Postbases fill position 3; positions 4-10 encode the seven template-ordered Pre-inflectional positions described in §3.2.2; positions 11-16 are for the formatives that mark the Verb Inflection, divided into Mood (11-12, where only 12 is obligatory) and person and number marking for S, O, and usually A. Finally, 17-20 are for Enclitics, which as noted follow a specific order and probably

⁶That is, in deciding what counts as an ‘element’ to fill planar structure positions, I'm taking Verb Bases and Postbases to be elements—fitting positions 2 and 3 respectively, even though they are often made up of smaller (but not fully compositional) pieces. This amounts to a decision to set as the threshold for planar analysis at a level somewhat above that of the smallest morphological formatives. A similar move for English might take the phrase *they are overwhelmed us* and consider *overwhelm* as the verbal core, even though it is a derived stem consisting of the prefix *over-* and the verb root *whelm*. I later call this the setting of a LEXEMIC THRESHOLD for purposes of analysis.

never number more than four at a time. The periphery surrounding 2-20 consists of zones 1 and 21, fore and aft. respectively.

The span 2-16 thus represents the traditional Verb Word, and is the orthographic word for many native speaking writers. The span 2-20 represents the traditional (Verb Word-hosted) Clitic Group. Some native speaking writers follow the orthography's convention of attaching each clitic to the word with a hyphen, while other writers use no hyphen, in effect taking 2-20 as the domain of the orthographic word (whereas it's not common at all to see the clitics written apart, as separate words).

On this analysis, most positions are slots because they can only be occupied by a single element at a time, not several, as spelled out explicitly for positions 4 through 20. Zones are assumed for the flanking positions 1 and 21 in order to accommodate multiple phrases and their components: this is simply a convenience in order to focus on grammar around the verb core (position 2). Internally, only position 3 is a zone since it can include sequences of zero or more Postbases whose order is strictly based on its scope over the verb. Finally, three distinct sequences seem to share a function: 11-12 for Mood; 13-16 for Person and Number of A, S, and O; and 17-20 for various mostly adverb-like enclitics.

As illustration, (21) is a biclausal phrase with two Verb Bases and therefore must be treated as involving two instances or parses of the verbal planar structure, the first (here abbreviated '1 {v}') focused on the Verb Base *aper-* 'to utter', which thus fills position 2 as verb core; and the second ('2 {v}') focused on the stative Verb Base *cuka-* 'to be fast' so that it fills position 2 as verb core for the second parse:

- (21) *ap* *-tu* *-llini* *-aq* *-a* *-a* *=llu =gguq*
1{v}: 2 -4 -9 -10 -12 -16 =18 =19
2{v}: [1 - - - - - - -]
 utter -always -evidently -would -IND -3SG.A+3SG.OBJ =& =QUOT
cuka *-u* *-na* *-ku*
[21 - - -]
2 -7 -12 -16
be.fast -lack -APPOS -3SG.OBJ
'And, it is said, s/he would always utter it slowly.'

It cannot be emphasized enough that, although the planar structure can be applied to transcribed Cup'ik clauses, it does not constitute a perspicuous description of those clauses, much less an account of how clauses are constructed in

Table 1: Planar structure anchored on the Cup'ik verb base. (Obligatory positions are bolded.)

Pos	Type	Elements
(1)	Zone	Nonverb (If nominal, Any grammatical function/Case; otherwise particle) [OPT]
(2)	Slot	Verb core: Lexemic Verb Base or minimal Verb Base [OBLIG]
(3)	Zone	VV Postbase [OPT]
(4)	Slot	ASPECT: <i>-tu-</i> 'always', <i>-yuite-</i> 'never', <i>-qar-</i> 'momentarily' [OPT]
(5)	Slot	REALIZATION1: <i>-yaaqe-</i> 'in vain', <i>-ngate-</i> 'seem', <i>-ksaite-</i> 'not yet' [OPT]
(6)	Slot	TENSE: <i>-llru-</i> 'did', <i>-cique-</i> 'will', <i>-ngaite-</i> 'won't' [OPT]
(7)	Slot	STATUS: <i>-nrite-</i> 'not'; <i>-yugnarqe-</i> 'may'; <i>-yugnaite-</i> 'definitely won't' [OPT]
(8)	Slot	REALIZATION2: <i>-yaaqe-</i> 'in vain', <i>-ngate-</i> 'seem' [OPT]
(9)	Slot	EVIDENTIAL: <i>-llini-</i> 'evidently'; <i>-lli-</i> 'perhaps' [OPT]
(10)	Slot	TENSE-MODAL: <i>-(g)aqe-</i> 'would have', <i>-ki-</i> (requires Optative) 'will; did (narrative reading)'; <i>-lqe-</i> (requires Indicative transitives) 'did' [OPT]
(11)	Slot	Mood [OPT]
(12)	Slot	Mood [OBLIG]
(13)	Slot	Person+number [OPT]
(14)	Slot	Person+number [OPT]
(15)	Slot	Person+number [OPT]
(16)	Slot	Person+number [OBLIG]
(17)	Slot	Enclitic [OPT]
(18)	Slot	Enclitic [OPT]
(19)	Slot	Enclitic [OPT]
(20)	Slot	Enclitic [OPT]
(21)	Zone	Nonverb (If nominal, Any GF/case otherwise particle) [OPT]

Cup'ik. If the planar structure (or the program of which it is a part) were a grammar, it would be a finite-state grammar, inching from one position to the next, linearly, without systematically incorporating recursion (other than of the trivial X^* sort of which finite-state grammars are capable). In contrast, the traditional 'own-terms' account given in §3 is equivalent to a context-free phrase-structure grammar that is recursive on many levels (even if filtered in certain ways, see §3.2.2), and captures how Postbases normally both select, and have scope over, the building Base. And yet, admittedly, even that doesn't capture certain nuances, e.g., that the two Enclitics in (21) have scope over the whole utterance and not just the Word that hosts them. Of course, as noted the planar structure has the advantage of allowing a simple kind of measurement of constituent classes (independent of their level of embedding, and across possibly distinct constituent-forming strategies), as well as simple cross-linguistic comparison. Both these feats would be challenging if some type of context-free phrase structure grammar were used as the basis for comparison.

5 Constituency diagnostics applied to the Cup'ik clause

We now apply CONSTITUENCY DIAGNOSTICS (also called constituency tests) taken, except as noted, from Tallman (2021). The main focus of this section is a description of the results of constituency diagnostics applied to Cup'ik over the planar structure in Table 1. By a constituency diagnostic we refer to some generalization over the constructions of the language that identifies a subspan in the planar structure. The following tests will be applied:

- Free occurrence
- Non-interruptability
- Repair domain
- Non-permutability
- Ciscategorical selection
- Subspan repetition/subspan selection
- Phonological domains: Prosodic
- Phonological domains: Segmental
- Biuniqueness deviation domains

Furthermore, where applicable, each test will be FRACTURED into subtests where criteria are defined more specifically, usually (but not always) with the result that one subtest can be termed MINIMAL and the other MAXIMAL, based on the length of the span it ends up identifying.

5.1 Free occurrence (2-16, 2-20)

FREE OCCURRENCE is defined as “[a] well-defined contiguous subspan of positions whose elements can be uttered as a minimal free form” (Tallman 2021: 16). This picks out the span 2-16, as shown in (22), which show that the three obligatory positions in the planar structure, 2, 12, and 16 must all three be present to constitute a minimal free form:

- (22) a. *tekit-u-t*
2-12-16
arrive-IND-3PL.S
‘they arrive(d)’
b. **Tekit-u*
2-12
c. *(t)*u-t*
12-16

But if we define the free span as containing one and only one free element (i.e., a maximal minimum free form) then the span can add enclitic positions, even though enclitics cannot occur by themselves (23). In this respect they are different from Particles, which can occur as free forms (24):

- (23) a. *tekit-u-t=am*
2-12-16=20
arrive-IND-3PL.S=but
‘but they arrive(d)’
b. *=am
=20

- (24) a. *tekit-u-t* *cali*
2-12-16 21
arrive-IND-3PL.S also
‘they arrive(d) also’

b. *Cali.*

21

also

‘Also, more.’

The two spans, 2-16 and 2-20 correspond, respectively, to the Word and the Clitic Group identified in §3.

5.2 (Non)-interruptability (2-16)

(NON)-INTERRUPTABILITY is defined as “[a] well-defined contiguous subspan of positions whose elements cannot be interrupted by elements of class I”, where ‘class I’ is some sort of ‘interrupting element’, whether a free form as defined by free occurrence, or some other test-definable subspan, that is to say, by some other span showing constituent properties (Tallman 2021: 16). For Cup’ik, the result is always the Verb Word (span 2-16), whether the interrupting element is a single free form or several free forms. Thus (25a) shows a well-formed span 2-16 while (25b) interposes *ukut* (*cuut*) ‘these (people)’ between the Verb Base and a semantically somewhat verb-like Postbase *-qatar-* ‘going to’:

- (25) a. *Qanrute-qatar-a-n-ka* *u-kut* (*cuu-t*)
 2-3-12-15-16 [21–] ([21–])
 talk.to-gonna-IND-3PL.OBJ-1SG.A this-ABS.PL person-ABS.PL
 ‘I’m gonna talk to them, these (people).’
 b. **Qanrute uku-t (cuu-t) qatar-a-n-ka*
 2 [21–] ([21–]) 3-12-15-16
 Intended reading: ‘Talk to these (people), I’m going to (do so) to them.’

Crucially, when enclitics are present, forming a Clitic Group, span 2-20, interruption after position 16 is optionally possible, by either one or several free forms. Then, the enclitics cease to be associated with the anchoring Verb Word for purposes of the constituency diagnostics and scope effects we are considering, and are instead associated with whatever is the last of the interrupting free Words (as in (26b), where Nominal Word(s) interrupt):

- (26) a. *Qanrute-qatar-a-n-ka=am* *u-kut* (*cuu-t*)
 2-3-12-15-16=20 [21–] ([21–])
 talk.to-gonna-IND-3PL.OBJ-1SG.A=but this-ABS.PL person-ABS.PL
 ‘But I’m gonna talk to them, these (people).’

- b. *Qanrute-qatar-a-n-ka* *u-kut* (*cuu-t*)=*am*
 2-3-12-15-16 [21-] ([21-])=20
 talk.to-gonna-IND-3PL.OBJ-1SG.A this-ABS.PL person-ABS.PL
 ‘But I’m gonna talk to them, these (people).’

The free form(s) *ukut* ‘these’ or *ukut cuut* ‘these people’ interrupt at the boundary between the inflection and the enclitic, in effect taking it as its own enclitic and forming a new Clitic Group with it. Thus, the Verb Word 2-16 is non-interruptable but the Clitic Group 2-20 is not non-interruptable.

There is, however, a periphrastic construction in which a Verb Base that contains one or more productive VV Postbases can be broken into two Verb Words. It might be taken as a way to interrupt the original Verb Word but it is different from what we’ve considered up to now in that the interrupting material isn’t, by any test, a constituent in its own right. A simple example is shown in (27), based on (25a), where the interrupting material is marked off with square brackets for clarity:

- (27) *Qanru[-llu-ki* (*u-kut* (*cuu-t*))
 2[-12-16 ([21-]) ([21-])
 talk.to[-APPOS-3PL.OBJ (this-ABS.PL (people-ABS.PL))
pi]-qatar-a-n-ka.
 2]-3-12-15-16
 do.so]-gonna-IND-3PL.OBJ-1SG.A
 ‘Talking to them/these ones/these people, I’m gonna do so to them’ = ‘I’m gonna talk to them/these ones/these people.’

Here the Verb Base *qanru(te)*- ‘to talk to’ is “finished off” by getting Appositional mood inflection (since it cannot stand alone without any inflection); and the original sequence is then resumed by mounting the VV Postbase *-qatar-* ‘to be going to V’ on the virtually empty, prothetic, implicitly anaphoric Verb Base *pi-* ‘to do V’, which then hosts not only *-qatar-* but the subsequent Indicative Mood inflection. In effect then, you get two complete run-throughs of the planar structure.

Interpretation of this pattern depends crucially on whether you consider the interrupting stretch, *-lluki (ukut (cuut)) pi-*, marked in the position analysis with square brackets, as a legitimate interrupting element. It certainly is not a constituent; rather, it works to fission the 2-16 (Verb Word) span in (25a) into two distinct and grammatically connected 2-16 spans, and in that sense might be seen

as strong confirmation of the 2-16 span. But it also works as evidence of a constituent break—in this case—between the Verb base *qanru(te)*- ‘talk to’ and the VV Postbase *-qatar-* ‘to be going to V’.⁷

5.3 Repair domain (2-16; 2-20)

I define REPAIR DOMAIN as the minimal domain observed by speakers when repairing an utterance; that is, the minimal planar distance backwards that a speaker must go if they make an error somewhere along the way and wish to start again. Although it is not one of the tests in Tallman (2021), it is somewhat related to free occurrence and to (non-)interruptability in that it gives evidence of what might be considered a minimal unit of production. Likewise, because repetition is involved, it is also related to subspan repetition.

In Cup'ik oral discourse, if a pause or speech error occurs in the midst of a span 2-16, the speaker will go all the way back to the beginning of the Word and start again, as shown in the following recorded text examples, where ‘–’ indicates self-interruption before repair:

- (28) a. *elliriq-lu-ni–* *elliriq-lu-tek*
 2-12-16- 2-12-16
 feel.orphaned-APPOS-3.REFL.SG.S– feel.orphaned-APPOS-3.REFL.DU.S
 ‘he, feeling orphaned– those.two, feeling orphaned’
- b. *akuyut-li-a–*
 2-9-12.16–
 paste-maybe-IND.3SG.A.3SG.OBJ–
 akuyu-qura-lli-a
 2-3-9-12.16
 paste-keep-maybe-IND.3SG.A.3SG.OBJ–
 ‘maybe she pasted it– maybe she kept pasting it.’
- c. *tuqu– tuqu-ne–* *tuqu-t-niara-llini-lu-ku*
 2– 2-6– 2-3-6-9-12-16
 die– die-will.soon– die-have-will.soon-evidently-APPOS-3SG.OBJ
 ‘die– will soon die– (she) will evidently soon have it die.’

⁷See Miyaoka 2012: 938 and Woodbury 2017: 545–546 for further discussion of this construction. The problem is that *-qatar-* ‘gonna’ here is somewhat verb-like, having immediate scope over *pi-* ‘do thusly’ but—because *pi-* is essentially a pro-verb bound by *qanrute-* ‘to talk to,’ *-qatar-* ‘gonna’ also has implicit scope over *qanrute-* ‘to talk to’ or (with inflection) *qanrulluki* ‘talking to them’. I return to the issue of VV Postbases having verb-like properties in §6.

- d. [tekiy-ngait– [tekiy-ngait]-ni-lu-ni
 [2-6– [2-6-]-3-12-16
 arrive-won’t– arrive-won’t-say-APPOS-3.REFL.SG.S
 ‘won’t arrive– saying he (himself) won’t arrive.’

(28a) is the simplest example, where a whole span 2-16 is uttered, but on realizing the speaker uttered singular *-ni* in position 16 when dual *-tek* was intended, he repeats the entire span verbatim, but with position 16 corrected. (28b) is also a case where a whole span 2-16 is uttered and then redone—but in that case, the repeat version adds an extra Postbase at position 3, leaving everything else the same. (28c) and (28d) show interruptions at various points within the 2-16 span, returning to repeat the position 2 verb core until a complete and satisfactory 2-16 span is reached. (28d) is of special interest because it involves a break after what might be considered an internal constituent (marked with square brackets and discussed at length in §5.6.3); nevertheless that constituent break is not sufficient to prevent cycling back to the beginning of the whole Word.

I have not knowingly encountered examples where a speaker picks up again somewhere in the middle of the span 2-16 without returning at least to position 2; nor do I find any examples of a self-interruption or self-correction beginning with only an Enclitic (positions 17-20). However, I have no very clear examples parallel to (28a)-(28d) where an Enclitic is repaired. On the strength of that negative evidence, then, we might consider span 2-20 as a weakly-supported maximal minimal repair domain and the span 2-16 as a much better-supported minimal minimal repair domain.

5.4 (Non)-permutability (2-16; 2-20)

(NON)-PERMUTABILITY is defined as “[a] well-defined contiguous subspan of positions that cannot be variably ordered with one another (if a-b, then b-a must not occur)”. It then is fractured into STRICT NON-PERMUTABILITY, where the elements always occur in a fixed order; vs. FLEXIBLE NON-PERMUTABILITY WITH SCOPE, where in addition to fixed order, it is possible for elements to order variably if there are differences in scope. (Tallman 2021: 16, 24).

The account in §3 and the planar structure in Table 1 both imply strict non-permutability. The positions, as far as we have seen, occur in the order given and not other orders, as shown in (29a) and (29b).

- (29) a. *im-na tekite-qata-llini-u-q=ggur=am*
 [1-] 2-3-9-12-16=19=20
 that.one-ABS.SG arrive-gonna-evidently-IND-3SG.s=it.is.said=but
 ‘But that person evidently is going to arrive.’
- b. **im-na qata-tekite-llini-u-q=ggur=am* (1-1 3-2-9-12-16=19=20)
 **im-na tekite-llini-qatar-tu-r=ggur=am* (1-1 2-9-3-12-16=19=20)
 **im-na tekite-qatar-tu-llini-r=ggur=am* (1-1 2-3-12-9-16=19=20)
 **im-na tekite-qata-llini-r-tu=ggur=am* (1-1 2-3-9-16-12=19=20)
 **im-na tekite-qata-llini-u=ggur=am-eq* (1-1 2-3-9-12=19=20-16)
 **im-na tekite-qata-llini-u-r=am=ggug* (1-1 2-3-9-12-16=20=19)
- c. *im-na=ggur=am tekite-qata-llini-u-q* (1-1=1 9-12-16)
im-na=ggug tekite-qata-llini-u-r=am (1-1=1 9-12-16=20)
im-na=am tekite-qata-llini-u-q=ggug (1-1=1 9-12-16=19)

By itself, (29a)-(29b) suggests strict non-permutability within the span 2-20, our maximal minimum free form (or Clitic Group); but (29c) adds a further wrinkle, showing that it is always possible to host a sentential-scope clitic like =*am* ‘but’ or =*ggug* ‘it is said’ on a different Word entirely, without any change in scopal effect; in that expanded light, as Natalie Weber (p.c.) pointed out to me, the relevant span may be only 2-16, our minimum minimal free form (or Word).

The next question however is whether either 2-20 or 2-16 shows strict non-permutability and not flexible non-permutability with scope. The formulation of the planar structure itself lacks strictness in a few places, most clearly since as we already saw in §3.2.2, positions 5 and 8 are alternative positions for (at least) two Postbases expressing categories of Realization(‘in vain’ and ‘seem’, see Table 1); and moreover, the choice of position (5 vs. 8) affects their scope in keeping with the Postbase Scope Rule (11). We saw this in (17), which I repeat as (30) using the planar structure position numbers and treating the complex Verb Base *naptar-te-* (Nominal Base *naptar-* ‘whitefish’ + NV Postbase *-te-* ‘to catch’) as the unsegmented occupant of position 2:

- (30) (=17)
- a. *Naptarc-aaqe-llru-u-nga*
 2-5-6-12-16
 whitefish.catch-in.vain-did-IND-1SG.s
 ‘I caught a whitefish, but alas.’, e.g., it got away

b. *naptarte-llru-yaaq-u-a*

2-6-8-12-16

whitefish.catch-did-in.vain-IND-1SG.S

‘I did alas catch some whitefish.’, e.g., a veiled offer

In (30)=(17), the VV Postbase *-yaaqe-* ‘in vain’ turns up in position 5 in (30a) and position 8 in (30b), on either side of position 6 *-llru-* ‘did’. Furthermore, ordering is free within position 3, a zone in which most VV Postbases occur, but are subject to the Postbase Scope Rule (11), so that any Postbase in position 3 will have scope over the Verb Base and any earlier position 3 Postbases (as will be discussed further below in §5.6).

In all then, we can characterize the whole span 2–20, when considered on its own, as showing flexible non-permutability with scope, but the flexibility is only due to the ordering of elements within zone position 3 and in certain positions in the span 5–8. Otherwise, the span 2-20 shows strict non-permutability. Furthermore, when we consider the mobility of clitics in a wider clausal context—as in (29c)—we find evidence that clitics (positions 17-20) are themselves permutable as long as they are hosted by a different whole Word, and we can in that context consider that only the span 2-16 shows strict non-permutability.

5.5 Ciscategorical selection (2-16; 4-16)

CISCATEGORIAL SELECTION is defined as “[A] well-defined contiguous subspan of positions whose elements can only semantically combine with one part of speech class” (Tallman 2021: 16). It then is fractured into MINIMAL (all in span are ciscategorical) vs. MAXIMAL (all outside the span are transcategorical) ciscategorical selection.

As a first approximation, the Cup’ik span 2-16, the Verb Word, should seem to demonstrate minimal ciscategorical selection given that both Inflection (by rule (6)) and Postbase derivation (by rule (8)) select specific categories. In contrast Enclitics, occupying the span 17-20, can be appended not just to Verb Words but to any full Word (by rule (18)), and must therefore be regarded as transcategorical:

(31) a. Verb Word + Enclitic

tekit-u-t=llu

2-12-16=18

arrive-IND-3PL.S=&

‘and they arrived.’

- b. Nominal Word + Enclitic

tau-na=llu

this-ABS.SG=&

‘and this.’

- c. Particle Words + Enclitic

ernerpak=llu

all.day=&

‘and all day ...’

In fact though, even within the span 2-16 there are exceptions to minimal ciscategorial selection: certain Postbases in position 3 turn up as both NN and VV Postbases and in that sense may be considered transcategorial because they select either Nominals or Verbs. These Postbases typically act semantically as modifiers (rather than heads) with respect to the Nominal Base or Verb Base; and with Verb Bases, they can even serve as modifiers not to the Verb Base itself, but to one of the Verb Base’s core arguments (discussed in Woodbury 2017):

- (32) a. *mikelngu-urluu-t*

child-poor-ABS.PL

‘the poor children.’

- b. *tekit-urlur-tu-t*

2-3-12-16

arrive-poor-IND-3PL.S

‘they poorly arrived; the poor ones arrived.’

It also might be ventured that within positions 13-16, person and number inflection, there is some transcategoriality since there is some etymological overlap in the person marking on nouns, and that on verbs in certain moods (as seen for the plural *-t* in (32a)-(32b)); but this is sporadic, irregular, and non-biunique to a degree that it is marginal and probably not a part of synchronic grammar.

We are left to conclude that, in light of these few transcategorial position 3 Postbases, it cannot strictly be the case that 2-16 is minimally ciscategorial. But it can be considered maximally ciscategorial, since other than in position 3, it is virtually totally ciscategorial, whereas as we have seen, the Enclitic positions beyond it, 17-20, are entirely transcategorial.

One final note before leaving ciscategoriality. It can be noted that the span 4-16, taken by itself, is genuinely minimally ciscategorial since all select (lexemic or derived) Verb Bases. Granted, 4-16 is out of range of our testing program,

which only looks at spans that include the verbal core. Nevertheless, if contiguous ciscategoriality is in some sense a sign of constituency, then it identifies as some sort of a constituent what earlier was termed the Pre-Inflection (span 4-10) in combination with the Verb Inflection (span 11-16). I return to the issue of possible beyond-verbal-core constituency in §6.

5.6 Subspan repetition/Subspan selection (2-3, 2-4, 2-7, 2-16)

SUBSPAN REPETITION is defined as ‘a well-defined contiguous subspan of positions that occurs more than once for a given construction’ (Tallman 2021: 16).

Many aspects of Cup’ik morphology/syntax fit here, including syntactic coordination/subordination, e.g., examples (21) and (39), and they would straightforwardly support the span 2-16 (the traditional Verb Word, Base through Inflection).

However, I wish to limit my consideration to what might be considered as a subtest or a related test that focuses especially on the SUBSPAN SELECTION properties of some of the hundreds of Postbases belonging to the zone position 3, specifically ones which lead to the repetition of certain alternative but interesting subspans. All involve subspans selected and scopally embedded by position 3 VV Postbases, leading to repetition insofar as the embedded material contains positions that overlap those of the embedding position 3 Postbases and any material that follows it. Specifically, there are four kinds of cases:

- Subspan selection by ordinary position 3 VV Postbases: Span 2-3
- Subspan selection by special position 3 VV Postbases: Span 2-4
- Subspan selection by special position 3 VV Postbases: Span 2-7
- Subspan selection by a special position 3 VV Postbase: ‘say X’: Span 2-16

By SELECTION I mean that the position 3 VV Postbase selects a span that is grammatically a (derived) Verb Base (as defined in (8)) or Verb Word, beginning with position 2 and ending somewhere later in the planar structure; and that the selecting VV Postbase has semantic scope over the selected subspan in keeping with the Postbase Scope Rule (11). The subspan involves repetition in all but the first of these four cases because it includes positions beyond position 3. I consider each in a separate subsection.

Although these selection patterns lead to some subspan repetition, it is not repetition in which *each* instance of the repeating subspan contains a position 2 verb core; or even where one instance of the repeating subspan implies a gapped

or missing verb core.⁸ Rather, it is the selecting VV Postbase (from any of the four subclasses just identified) which does the selecting and seems to behave in certain respects like a verb core; but on the program being applied is prevented from being measured as a verb core because it is a suffix and not a root. Accordingly, it might be better to consider subspan selection as different test from subspan repetition, because subspan selection focuses on the embedded span alone, rather than the overlap between the embedding and the embedded span.⁹ This is entirely an artifact of our method, one that becomes conspicuous as the method tries to measure Cup'ik grammar: namely, our method does not recognize the VV Postbases in question as verbal cores in their own right—if it did, then what we have to call subspan selection would be subspan repetition in the notionally intended sense. We return to this question in §6. But for our purposes, subspan selection is useful and straightforward in demonstrating a certain type of constituency.

5.6.1 Subspan selection by ordinary position 3 VV Postbases: Span 2-3

Most VV Postbases occur in position 3, a zone, defined in traditional terms as a VV Postbase applying via the Base Recursion Rule (8) to a lexemic Verb Base (occupying only position 2) or to a complex Verb Base already derived via (8). The distinguishing feature of position 3 VV Postbases is that they cannot follow a Postbase of position 4 or later and they normally obey the Postbase Scope Rule (11), having scope over the Verb Base they select:

- (33) *Prime-at=ll' taw' an-te-qata-llini-lu-ku*
1-1=1 1 2-3-3-9-12-16
prime-ABS.PL=& then go.out-with-gonna-evidently-APPOS-3SG.OBJ
'And the Primes (basketball team) are apparently about to take it (the ball) out (into the court).'

In (33), position 3 *-(u)te-* 'V along with' selects and has scope over position 2 *an(e)-* 'go out' and *-qata(r)-* 'gonna V' selects and has scope over the 2-3 span *an-te-* 'to take out'. Then *-llini-* 'evidently' selects and has scope over the 2-3 span *an-te-qata(r)-* 'to be going to take out'. That is followed by the Appositional Mood, and a person/number ending. But none of (34) are well formed:

⁸That is, suppose a position 3 Postbase selects a subspan 2-7 (the case considered in §5.6.3). Then, the selected span might, say, consist of 2-3-7; and the selecting Postbase might be followed by the sequence 7-12-16. In that case the only actually repeating sequence will be 3-7, and it will not include the verb core itself.

⁹I thank Natalie Weber for a really perspicuous discussion of this difference.

- (34) **an-te-llini-qatar-lu-ku* (2-3-9-3-12-16)
**an-llini-te-qatar-lu-ku* (2-9-3-3-12-16)
**an-qata-ut-llini-lu-ku* (2-3-3-10-12-16)

Although the last of these does not violate the planar structure, it apparently does not make sense for position 3 –(*u*)*te*- ‘V along with’ to select and have scope over the (otherwise well-formed) 2-3 span *an-qata(r)*- ‘to be going to go out’.

5.6.2 Subspan selection by special position 3 VV Postbases: Span 2-4

Certain VV Postbases select a subspan from position 2 to a position after position 3. All these Postbases add a transitive subject argument and have meanings like ‘cause’, ‘say’, ‘tell’, ‘wait for’, and more. They break down into three sets according to the span they select, to be treated, respectively, in this and the following two subsections.

The first set selects the subspan 2-4. Its members have been termed (among other things) DOUBLE TRANSITIVE (Kleinschmidt 1851), and include the following VV Postbases (Woodbury 2005):

- ‘want, ask, tell (subject) to V
- ‘think that (subject) might V
- *-cite-* ~ *-vkar-* (suppletive) ‘let, allow, cause (subject) to V’

The examples in (35), involving *-sqe-* ‘want, ask tell (subject) to V’, illustrate:

- (35) a. [*Kinerci-qaa*]-*sqe-vke-na-ki*
[2-4]-3-7-12¹⁰
[dry.something-just]-tell-not-APPOS-3PL.OBJ
‘Telling (him) not to just dry them.’ (Literally: ‘Not telling (him) to just dry them’)
- b. **[Kinerci-qa-nrite]-sqe-llu-ki*
[2-4-7]-3-12-16
Intended reading: ‘Telling him not to just dry them.’

¹⁰Strictly speaking, this is a somewhat illegal planar representation since ‘3’ follows ‘4’; but it offers a simple and perspicuous way to capture the embedding and recursion of subspan selection without positing a great many more positions only to capture the workings of this (and the following) limited set of Postbases.

In these examples, brackets enclose the subspan selected by the position 3 VV Postbase *-sqe-*. In (35a), *-sqe-* selects a subspan ending in the position 4 VV Postbase *-qar-* ‘just V’; but it cannot select a subspan ending with the position 7 negative VV Postbase (*-nrte-* ‘not’).¹¹

5.6.3 Subspan selection by special position 3 VV Postbases: Span 2-7

The next set of position 3 VV Postbases is also a part of the Double Transitive class recognized by Kleinschmidt, but it has scope over the subspan 2-7 and includes these among others:

- ‘think that V’
- ‘say that V’

Here again, brackets mark the embedded subspan:

- (36) [qacingqa-nri]-**cu**k-lu-ki
[2-7]-3-12-16
[stay.put-not-]-think.that-APPOS-3PL.OBJ
'thinking they were not staying put.'
- (37) [atanqe-ciq]-**ni**-llru-a-teng
[2-6]-3-6-12-16
wait.for-will-say.that-did-CONSEQ-3SG.A.3REFL.PL.OBJ there-LOC
'because he said (he) will wait for them there.'

In (36), *-yuke-* ‘to think that V’ (appearing here as *-cuk-*) follows the position 7 VV Postbase *-nrite-* ‘to not V’; in (37), *-ni-* ‘to say that V’ follows the position 6 VV Postbase *-ciqe-* ‘will’, which applies to the waiting event; but then starts its own recursive “cycle” that allows it to be followed by its “own” position 6 VV Postbase indicating the tense of the whole event of saying, *-llru-* ‘did’.¹²

¹¹The negation markers *-vke-* in (35a) and *-nrite-* in (35b) are suppletive allomorphs in complementary distribution, the former occurring when the Appositional Mood immediately follows and the latter occurring elsewhere. Both examples happen to be in the Appositional Mood, which, as noted, does not overtly index the transitive subject (A). But they are present and recoverable, because normally—including in these particular cases—the implicit A is bound by a (transitive or intransitive) subject elsewhere in the discourse.

¹²Recall also the repair example (28d), where a *-ni-* ‘say’ embeds a span 2–6.

5.6.4 Subspan selection by a special position 3 Postbase ‘say X’: Span 2-16

The position 3 VV Postbase *-a(a)r-* ‘say Word (to someone)’ embeds a subspan 2-16—that is, any whole Word (including any Verb Word):

- (38) a. *Pik-a-qa-ar-lu-ku.*
[2-12-16-]-3-12-16
own-IND.3SG.OBJ-1SG.A-say-APPOS-3s.OBJ
‘Saying to him/her, “It’s mine!”’
- b. *Pik-a-qa!*
2-12-16
own-IND.3SG.OBJ-1SG.A
‘It’s mine!’ (Literally: ‘I have it as a thing, I own it.’)

In the traditional own-terms framework, this and similar constructions are taken as one of a few synchronic instances where a verbalizing Postbase is added to a Word to form a new complex Base (i.e., Base → Word + Postbase; see e.g., de Reuse 1994, Sadock 2017). One might ask whether the subspan in this case might instead be a result of compounding or cliticization; yet with respect to the other tests presented in this section, the whole sequence from position 2 to position 3 is a derived Verb Base just like any other, and the whole resulting Verb Word (as in (38a)) shows no phonological evidence of an enclitic boundary between the internal inflection and the verbalizing Postbase.

5.6.5 Conclusion

The upshot of this discussion is that position 3 VV Postbases break down into groups according to the subspans that they select and have scope over. For our purposes these suggest constituent properties for the relevant subspans: 2-3, 2-4, 2-7, and 2-16. Complementarily, they also strongly suggest constituent properties for a constituent anchored by the position 3 VV Postbase itself, which behaves in many ways like a lexical verb in its own right, even though it is not what we are calling the verb core of the construction. We will return to this question in §6.

5.7 Phonological domains (2-16; 2-20)

Tallman (2021: 16) defines three categories of phonological domain—segmental, stress, and tone where the application of phonological or morphophonological processes may define contiguous subspans as constituents of some kind. In this

section I will focus on two prominent sets of phonological phenomena, one prosodic and the other segmental.

5.7.1 Prosodic domains (2-16; 2-20)

The domains of (mostly) automatic prosodic foot and stress assignment rules in Cup'ik are ideal for our purposes since they are clear, dramatic, and well-studied. These rules define iambic feet, from left to right, beginning at the Base of any Word (therefore at position 2 for Verbs) and ending after the Inflection (position 16 for Verbs). The syllabic shape conditions and internal composition of feet will not concern us (see Woodbury 1981, 1987 for details). In (39) we have two Verb Bases, and for purposes of measurement, each can function as position 2 verb core *pissu-* 'to hunt' defining the first (labeled 1 {v}), and *mallussu-* 'to hunt beached whales' defining the second (labeled 2 {v}):

- (39)
- | | | | | | | | |
|----------|---------------|------------|---------------|------------|------------|--------------------|------------|
| | <i>pissu</i> | <i>-tu</i> | <i>-llini</i> | <i>-lu</i> | <i>-ni</i> | <i>mallussu</i> | <i>-tu</i> |
| 1{v}: 2 | -4 | -9 | -12 | -16 | [21 | - | |
| 2{v}: [1 | - | - | - | -] | 2 | -4 | |
| | hunt | -always | -evidently | -APPOS | -3SG.S | hunt.beached.whale | -always |
| | <i>-llini</i> | <i>-lu</i> | <i>-ni</i> | | | | |
| | - | - | -] | | | | |
| | -9 | -12 | -16 | | | | |
| | -evidently | -APPOS | -3SG.S | | | | |
- 'He apparently always hunting, he apparently always hunting beached whales.'

The footing rules apply to the span 2-16 in each parse (that is, they apply to the traditionally-recognized Word), where they group light syllables into iambic binary feet, stressing and (if open) lengthening each foot-final syllable; but the foot rules stop short of footing the final syllable of the span, therefore always leaving one (first parse) or else two (second parse) final syllables unfooted, unstressed, and unlengthened. We see this in the annotated phonetic form presented in (40a), where syllables are broken with periods and feet shown with parentheses. Crucially, the footing process cannot continue, unabated, from one Word into the next; it must reset and start again with each new Word, as shown by the impossibility of (40b):

- (40)
- a. →[(pi.'su:)(tu.'fi:)(ni.'lu:.)ni (ma.'tu:)(su.'tu:)(fi.'ni:.)lu.ni]
- b. * →[(pi.'su:)(tu.'fi:)(ni.'lu:)(ni.'ma:)(tu.'su:)(tu.'fi:)(ni.'lu:.)ni]

Iambic foot formation does, however, continue without reset through any Enclitics that may follow span 2-16; that is, through position 20. For example, if we add Enclitics =*llu* ‘and’ =*gguq* ‘it is said’ into positions 19 and 20, respectively, for each of the two 2-16 spans just considered, we get a continuation of the same foot formation pattern – shown in (41) using phonetic transcription – where iambs continue to be gathered until the penultimate syllable, leading, in the cases below, to one new foot each:

- (41) a. →[(pi.'su:.) (tu.'hi:.) (ni.'lu:.) (ni.'=tu:.) =xuq]
‘And, it is said, he apparently always hunting’
b. →[(ma.'tu:.) (su.'tu:.) (hi.'ni:.) (lu.'ni:.) =tu: =xuq]
‘And, it is said, he apparently always hunting beached whales’

Nevertheless, the feet created when Enclitics are included follow a slightly different pattern, not visible in the simple examples just shown (see Miyaoka 1985 and Woodbury 2002: 93 for the details of this). Therefore, it is best to say that 2-16 is the span for the core set of footing rules, while 2-20 is the span for a closely related by slightly adjusted further set of footing rules.

5.7.2 Segmental domains (2-16)

Extensive morphophonemic processes apply at the junctures within the span 2-16; whereas few if any of these processes apply at Enclitic junctures, positions 17-20 (Reed et al. 1977; Woodbury 1981; Miyaoka 2012). Crucially, these processes also do not apply across the juncture from position 1 to position 2, leading to virtually total morphophonological invariance for the onsets of Verb Bases. We now describe three such cases.

5.7.2.1 Syllabic structure and VVV cluster avoidance (2-16)

The phonemic representation of any Word, including the Verb Word (span 2-16), consists of one or more syllables with the shapes CV, CVC, CVV, or CVVC (where VV is any combination of the peripheral vowels /a, i, u/ but never the central vowel /ə/). Word-initially (and hence at the onset of any Verb Base), the initial syllable can have no consonant onset, hence V, VC, VV, and VVC are also allowed. These patterns are amply illustrated, and never counterexemplified, in the examples in this article. Likewise, Enclitics (span 17-20) can form syllables with no vowel onset, e.g. =*am* ‘but’. Furthermore, no phonemic representation of a span 2-16, or of any Enclitic, ever ends in the central vowel /ə/, serving phonologically to (negatively) demarcate the end of the spans 2-16 on up to 2-20.

One consequence of the foregoing is that it should never be possible to find a VVV sequence within the span 2-16; and this is true. But because Verb Bases (position 2) can begin with a vowel, it should be possible to find a VVV sequence across a span 1-2; and likewise anywhere after position 16. And it is:

- (42) a. *Qaill' ma-kut qulira-t pi-aq-ata*
 1 1-1 1-1 1-1-1
 how this-ABS.PL stories-ABS.PL tell-CONTIN-3PL.S
aanait-aq-u-t qaa?
 2-10-12-16 21
 lack.mother-would-IND-3PL.S Q
 'Why is it that whenever they tell these stories, there would be no mother?'
- b. *Cuuci-a assiiri-u-q*
 1-1 2-12-16
 life-ABS.SG+3SG.POSS worsen-IND-3SG.S
 'his life got worse.'
- c. *teki-ca-mi-u=am*
 2-12-15-16=20
 reach-CONSEQ-3REFL.SG.A-3SG.OBJ=but
 'but when he reached it.'
- d. *qumiu-llr-a-t-ni=llu*
 2-12-14-15-16=18
 be.pregnant-CONTMP-3SG.OBJ-3PL.A-CONTMP=and
aana-ita
 21-21
 mother-REL.PL+3PL.POSS
 'and while their mothers were pregnant.'

In (42a)-(42b), VVV sequences arise between positions 1 and 2 between two Verb Words; in (42c) between positions 15-16 and 20, that is a Verb Word and an Enclitic; and in (42d) between positions 18 and 21, an Enclitic hosted by a Verb Word and a following Nominal Word.

Finally, we can note that many morphophonological processes operating in the Cup'ik Word (including not only the Verb Word span 2-16, but also Nominal and Particle Words) lead to the total avoidance of VVV clusters through epenthesis, hiatus, and constraints against otherwise normal intervocalic consonant loss when it would lead to a VVV cluster: for details see Woodbury 1981: 29–103; Reed et al. 1977: 18–38; and Miyaoka 2012: 195–219.

5.7.2.2 Uvular-velar consonant coalescence (2-16)

The morphophonological processes just mentioned belong to an even larger suite of processes that (a) apply only within the Cup'ik Word (meaning, for Verb Words, the span 2-16 and excluding Enclitics (17-20); (b) are partly automatic, conditioned by the segments at the end of the building Base and the beginning of each new Postbase or Inflection that follows; and (c) are partly idiosyncratic, requiring that Bases, Postbases, and Inflectional suffixes must be grouped in morphophonological or morpholexical classes whose exact alternations and behavior cannot entirely be predicted by automatic morphophonological rules, but which still allow—once classes are carefully established—for powerful generalizations. There is not space to demonstrate very much of this here, but I have picked out a representative example demonstrating the three characteristics (a)-(c) just noted, a rule of UVULAR-VELAR CONSONANT COALESCENCE.

Bases and Postbases end either in a vowel or in a (non-labialized) velar or uvular continuant, usually /ɣ,ʁ/ but very rarely their voiceless counterparts. An arbitrary subgroup of /k/-initial Postbases and inflectional suffixes comprise a special class, such that when any member of that class is suffixed to a /ʁ/-final Base or Postbase, the resultant /ʁk/ cluster coalesces as /q/.¹³ We will consider four suffixes from this class, each occupying a different position. The suffixes are given in IPA-based morphophonemic representation where the coalescing /k/ morphophoneme is underlined to distinguish it from suffix-initial /k/ morphophonemes with other behaviors:

- /-ksaitə-/ 'to not yet V' (position 5)
- /-ki-/ 'will' (with Optative) (position 10)
- /-kə-/ Transitive Participle Mood (position 12)
- /-ka/ 1st Person Singular Transitive Subject (position 16)

We will start by considering them in combination with two position 2 Verb Bases, shown here in morphophonemic citation form and in the Appositional Mood (for third person singular direct object), which fully preserves all segments:

- /təɣu-/ 'to take' (cf. /təɣu-lu-ku/ 'taking it')
- /atuʁ-/ 'to use' (cf. /atuʁ-lu-ku/ 'using it')

¹³It is also the case that when a /ʁk/ cluster involving this class arises, the cluster simplifies to just /k/; but for our purposes that can be left aside.

/təyu-/ 'to take', is vowel-final and should thus not affect the suffix initial morphophoneme /k/, whereas /atuɤ-/ 'to use' is /ɤ/-final and thus it should show coalescence. This can be seen for the first three suffixes in (43)-(45):

- (43) a. /təyu-ksait-a-a/
 2-5-12-16
 take-not.yet-IND-3SG.A+3SG.OBJ
 's/he has not taken it yet.'
- b. /atu-qsait-a-a/
 2-5-12-16
 use-not.yet-IND.3SG.A+3SG.OBJ
 's/he has not used it yet.'
- (44) a. /təyu-ki-li-u/
 2-10-12-16
 take-will-OPT.3SG.A-3SG.OBJ
 's/he should take it.'
- b. /atu-qi-li-u/
 2-10-12-16
 use-will-OPT.3SG.A-3SG.OBJ
 's/he should use it.'
- (45) a. /təyu-kə-ka/
 2-12-16
 take-TRPRT.3SG.OBJ-1SG.A
 'that I take it.'
- b. /atu-qə-ka/
 2-12-16
 use-TRPRT.3SG.OBJ-1SG.A
 'that I use it.'

It will be noticed that the examples in (45) also show the fourth /k̥/-coalescing suffix in position 16, the first person singular transitive subject suffix /-k̥a/. There it occurs after the final vowel of the Transitive Participle /-kə-/ and for that reason undergoes no coalescence. But (46) shows that same suffix after the Indicative suffix /-aɤ-/ in (46a), where coalescence indeed occurs; meanwhile, (46b) demonstrates a context where the /ɤ/ that is part of the Indicative is present:

- (46) a. /təyu-a-qa/
 2-12-16
 take-IND.3SG.OBJ-1SG.A
 ‘I take it.’
 b. /təyu-aʁ-put/
 2-12-16
 use-IND.3SG.OBJ-1PL.A
 ‘we use it.’

(43–46) then show that certain /k/-initial formatives show coalescence throughout the 2-16 span. Meanwhile no /k/-initial position 2 Verb Base or position 17-20 Enclitics coalesce with preceding uvulars, ever; instead, uvular-/k/ sequences are preserved, as shown for a /k/-initial Verb Base in (47a) and (the only) /k/-initial Enclitic in (47b):¹⁴

- (47) a. /nanvaq kau-lu-ku/
 1 2-12-16
 lake.ABS.SG reach.into-APPOS-3SG.OBJ
 ‘reaching into the lake.’
 b. /tʃa-niar-tu-q=kiq/
 2-10-12-16=19
 do.what-should-IND-3S=I.wonder=REFL.
 ‘I wonder what he should do?’

In summary then, uvular-velar consonant coalescence in Cup’ik (a) operates across formative boundaries only in the span 2-16; (b) it is a more-or-less natural phonological process; and yet (c) it is idiosyncratic in that not all /k/-initial suffixes behave this way. Citing the technical literature on Central Alaskan Yupik morphophonemics, I have claimed that these characteristics are broadly characteristic of the language and lead to the variability that you see when phonemic-level forms are segmented (see Footnote 1).

¹⁴It will be noted that the uvular plus /k/ sequences involve the stop /q/ and not the continuant /ɣ/, as in prior examples. This actually reflects a further Word (including Verb Word, span 2-16) privilege, namely that formative-final velars and uvulars are only continuants within 2-16 (if they appear at all), and only stops elsewhere, the situation for the /q/’s in (47a) and (47b). In fact in Nominal Bases, where the Absolutive Singular form has no overt inflectional suffix, Base-final /ɣ/ surfaces as /q/ when (notionally) Word-final but as /ɣ/ after Postbases that do not delete it: thus /nanvaq/ ‘lake’ in (47a) is /q/-final as an independent word but /ɣ/-final when followed by, say, the Posbase /-kaq/ ‘what will be N’: /nanvaʁ-kaq/ ‘what will be a lake’.

5.7.3 Phonological conclusions

It is of particular interest that the span claimed for segmental processes throughout the preceding section is 2–16, and not anything smaller. Given, for example, Lexical Phonology and Morphology (Kiparsky 1985) and other frameworks that have built on its insights, one might expect levels or layers radiating out from the lexical Verb Base, where spans like 2–3 (derived Verb Bases minus the Pre-Inflection) or 2–10 (all derivation, minus inflection) might be expected to show partly different morphophonological processes. But this has not been a finding, at least so far, of Central Alaskan Yupik morphophonological research. Rather, just as with uvular-velar consonant coalescence, morphophonological processes are distributed evenly throughout the whole span 2–16. The only significant layering, then, is when Enclitics (17–20) are added. They share with the span 2–16 the propagation of left-to-right iambic footing, as shown earlier; but they are otherwise like independent words in lacking segmental morphophonological processes.

5.8 Biuniqueness deviation domains (11–16; 4–12; none that includes verb core)

A BIUNIQUENESS DEVIATION DOMAIN is defined as “a well-defined contiguous subspan of positions whose elements display deviations from biuniqueness (one meaning-one form)” (Tallman 2021: 16). All things considered, Cup'ik is remarkably biunique, given its relatively fusional morphophonology, and yet there are two spans where biuniqueness sometimes breaks down, and interestingly, they do not involve the verb core. For that reason, these domains are, strictly speaking, not germane to the testing program at hand, which only considers constituent domains that include the verb core. Nevertheless I will discuss these extra-verb core domains in any case and then suggest later, in §6, how they may inform a somewhat differently-conceived exploration of constituency using planar methods.

5.8.1 Nonbiunique marking of mood and person/number of S/A/O: Span 11–16

Verb Inflection includes marking for Mood in positions 11–12 and for the person and number of S, O, and (with most Moods) A in positions 13–16. But within this span 11–16 there often is suppletion, zero marking, cumulative exponence, or multiple exponence, depending on the Mood and person/number combinations involved. For example:

- (48) a. *tangrr-ar-pe-kut*
2-12-14-16
see-IND-2SG.A-1PL.OBJ
'you (sg.) see us.'
- b. *tanger-Ø-kut*
2-12-16
see-OPT.2SG.A-1PL.OBJ
'(You, sg.) see us!'
- c. *tangerr-lu-ta*
2-12-16
see-APPOS-1PL.OBJ
'seeing us.'
- d. *tangrr-a-i*
2-12-16
see-IND-3SG.A+3PL.OBJ
's/he sees them.'
- e. *tangrr-a-g-ke-t*
2-12-13-15-16
see-IND-3DU.OBJ-3DU.OBJ-3PL.A
'they see those two.'
- f. *tangrr-aq-a-ne-g-ne-ki*
2-11-12-13-14-15-16
see-CNTG-CNTG-3DU.A-3DU.A-3DU.A-3PL.OBJ
'whenever those two see them.'

In (48a)-(48b) VS. (48c) the first person plural object marker shows suppletion (-*kut* ~ -*ta*); In (48b) there is zero marking for the Optative mood and the second person singular A subject, i.e., cumulative exponence; in (48d) there also is cumulative exponence, with -*i*- marking both A and O; in (48e) the ordering of A vs. O is opposite to that in (48a), and the third person dual object is marked twice, i.e., has multiple exponence, while the third person singular A subject is marked by -*t*, which is suppletive with respect to its marking with -*a* in (48d); and in (48f), there is extreme multiple exponence, with two formatives marking the Contingent Mood and three marking the third person dual A subject.

5.8.2 Cumulative exponence and suppletion involving negation: Span 4-12

The span 4-12 includes what we have called the Pre-Inflection (4-10) and Mood marking (11-12) within the Verb Inflection. Negation, which normally occurs in position 7, fuses or forms portmanteaux or induces suppletive forms for certain elements in neighboring positions when they are adjacent. The following are examples:

- (49) *-nrite-* 'not' (7) + *-lu-* 'Appositional mood' (12) → *vke-na-*
- a. *tegu-nrit-a-a*
2-7-12-16
take-not-IND+3SG.OBJ-3SG.A
's/he doesn't take it.'
 - b. *tegu-lu-ku*
2-12-16
take-APPOS-3SG.OBJ
'taking it.'
 - c. *tegu-vke-na-ku*
2-7-12-16
take-not-APPOS-3SG.OBJ
'not taking it.'
- (50) *-ciqe-* 'will V' (6) + *-nrite-* 'not V' (7) → *+ngaite-* 'will not V' (6)
- a. *an'e-ci-q-u-a*
2-6-12-16
go.out-will-IND-1SG.S
'I will go out.'
 - b. *an'e-nrite-u-a*
2-7-12-16
go.out-not-IND-1SG.S
'I am not going out.'
 - c. *an-ngait-u-a*
2-6-12-16
go.out-will.not-IND-1SG.S
'I will not go out.'
 - d. **an-ci-qe-nrit-u-a*
2-6-7-12-16
go.out-will-not-IND-1SG.S

- e. *ane-llru-nrit-u-a*
2-6-7-12-16
go.out-did-not-IND-1SG.S
'I did not go out.'

(51) *-tu-* 'always' (4) + *-nrite-* 'not V' (7) → *-yuite-* 'never' (4)

- a. *an'e-tu-u-nga*
2-4-12-16
go.out-always-IND-1SG.S
'I (always) go out.'
- b. *an-yuit-u-a*
2-4-12-16
go.out-never-IND-1SG.S
'I never go out.'
- c. **an-tu-nrit-u-a*
2-4-7-12-16
go.out-always-not-IND-1SG.S

In (49), the position 7 negative 'not' Postbase and the position 12 Appositional Mood suffix undergo mutual suppletion when (and only when) adjacent. In (50), the position 6 future 'will' Postbase and position 7 negative 'not' Postbase are obligatorily replaced with a suppletive portmanteau when adjacent, which, as shown in (50e) does not happen when the past-tense 'did' Postbase occupies Position 6 before negation in position 7. And in (51) the position 4 habitual 'always' Postbase and position 7 negation 'not' Postbase are obligatorily replaced with a suppletive portmanteau when adjacent.

Both sets of deviations from biuniqueness suggests "patches" of constituency. In the 11-16 case, that "patch" is clearly the Verb Inflection as a whole; and in the 4-12 cases with negation, it is the formation of something like an incipient negative auxiliary. Both are islands of "wordiness" that tend to exclude the verb core (position 2)–as well as the Postbases in zone 3. As such they are beyond the present project, which only considers spans that include the verb core. They nevertheless are potentially of interest when we take a more neutral view of constituency tests, as discussed in §6.

5.9 Summary and conclusion

Our constituency diagnostic results from this section are summarized in Table 2. In the following discussion, I make some basic generalizations over the results and compare them to the "own-terms" traditional analysis presented in §3.

Of the nine diagnostic types we can recognize, all but biuniqueness deviation support the traditional Verb Word (2–16) as a constituent, while three-free occurrence, non-permutability, and prosodic phonological domains–also support the Clitic Group (2–20). Note that both receive support among the first six, more morphosyntactically-based diagnostics; as well as among the two (morpho)phonologically-based diagnostics. It would be wrong, for example, to say that the Verb Word (2–16) is a grammatical construct only, since it is a major phonological domain; and equally wrong to say that the Clitic Group (2–20) is only a phonological construct, owing to its non-permutability with respect to a maximal fracturing.

Table 2: Constituency diagnostic results for Cup'ik (^a = excludes verb core.)

Constituency diagnostic	Min frac	Max frac
Free occurrence	2–16	2–20
Non-interruptability	2–16	2–16
Repair domain	2–16	--
Non-permutability (Flexible/scopal within 3 and 5–9; rigid otherwise)	2–16	2–20
Ciscategorical selection	4–16 ^a	2–16
Subspan repetition	2–3, 2–4, 2–7, 2–16	
Phonological domains: Prosodic	2–16	2–20
Phonological domains: Segmental	2–16	--
Biuniqueness deviation domains	4–12 ^a ; 11–16 ^a	

Sporadically, shorter subspans are also supported, but with no clear ‘winners’; and some of these are ‘illegal’ if we only consider constituency that includes the verb core. The (traditional) Verb Inflection is singled out as a site of biuniqueness deviation; but, somewhat surprisingly, its complement, the maximal Verb Base (2–10) is not, and that boundary is muddled further as biuniqueness deviation singles out 11–16, which groups the “own-terms” Sub-Inflection and Mood. The following summarizes:

- Span 2-3 (Verb Base without adjunction of Subinflection (4-10)), by subspan repetition.
- Span 2-4 (Verb Base including adjunction of Aspect), by subspan repetition.

- Span 2-7 (Verb Base including adjunction of Aspect, Tense and Negation), by subspan repetition.
- Spans 3 and 5–8, by non-permutability that is flexible with scope.
- Span 4–12, the combined subinflection (4–10) and Mood (11–12) spans when negation (normally position 7) is present, by biuniqueness deviation (suppletion and cumulative exponence)
- Span 11-16 (the whole Verb Inflection), by biuniqueness deviation (of all kinds)

It is also notable that, as indicated, none of these spans gets major support on purely phonological grounds. For example, the stress rules discussed in §5.7.1 cannot “see” internal divisions within the span 2-16; nor are there segmental morphophonological processes that pertain to subspans of 2-16. In general, we do get several levels of grammatical elaboration of the recursive, left-branching Verb Base (2-3, 2-4, 2–7). We also get a “patch” where early VV Postbases show scopal effects (within 3 and sporadically in 5-8 due to “wild cards” occupying positions 5 and 8 according to scope). Arguably, the true scopal domain—per the Postbase Scope Rule (11)—might be considered the span 2-4 and then, sporadically, 5-8 as noted. This is because the non-permutability of the verb core and 3-8 is scope-based. If so, then what we actually find is a cascading series of left-branching constituents, as predicted by the Base Recursion Rule (8).

But as noted, biuniqueness deviation shows us two “patches” of constituency behavior that lie beyond the verb core, 4-12 (Subinflection with Mood), and 11-16 (the Verb Inflection proper). These pose auxiliary-verb like clusters within the word distinct from the left-branching, recursive Verb Base, and may pose what may superficially be described as a ‘bracketing paradox.’

People – linguists and non-linguists alike – without much acquaintance with UYI grammar and phonology are often skeptical, asking, How could the words of a language be that long? Surely these are phrases written without spaces, and not actual words. But the cumulative weight of our constituency tests—mostly drawn from among those tests typically thought of as being diagnostic of wordhood (as reviewed in Haspelmath 2011 and Tallman 2020)—offers considerable ballast to the idea, accepted by UYI-family native speakers, and by Native and non-Native specialists in UYI linguistics, that these long stretches are indeed words, without much problem at all.

6 The “verb core”, and gauging holophrasis directly: theoretical and empirical issues

Let us turn attention back to our comparative program, which—as we just saw—offers strong support for the traditionally recognized Verb Word and Clitic Group. We also saw slivers of support for subspans within the Verb Word. But if, as contended, UYI languages are highly holophrastic, why only slivers? Are there some general ways to amend our program so that it more fully detects constituency within the Verb Word, or, more generally, constituency within the whole clause that might even dissect the Verb Word or reapportion its pieces, so as to offer a better basis for the holophrasis intuition?

Here my goal is to point out directions, rather than offer full solutions.

Consider this: Dixon & Aikhenvald (2002: 19–20), in characterizing the “grammatical word”, cite “conventionalized coherence and meaning”, specifying that “while the meaning of a word is related to the meanings of its parts, it is often not exactly inferable from them.” Because of such idiosyncrasy, words should be listable in the lexicon. But clearly, the best analog to “grammatical words” in this sense is not the whole (traditional) Verb Base (span 2-10, including all but the Verb Inflection) or even for that matter the span 2-3, that is, the Verb Base minus span 4-10, what we called the Templatic Pre-Inflection, since the recursiveness of productive Verb Base formation (8) is theoretically infinite. Rather, the best analogs to grammatical words by Dixon and Aikhenvald’s criterion are lexemic Verb Bases (position 2 only) and at least some VV Postbases, which, as we have seen, are productive, have conventionalized coherence and meaning, and, when internally complex and composed themselves of suffix pieces (as shown in §3.4.1), have meanings that aren’t always inferable from the meanings of those suffix pieces. In other words, Dixon and Aikhenvald’s criterion—especially taken together with our results in the previous section—give us exactly what we need to recognize the holophrasis we encounter in UYI languages like Cup’ik.

Consider again expressions like (36), repeated here as (52):

- (52) = (36)
 [*qacingqa-nri-*]-*cuk-lu-ki*
 [2–7]-3-12-16
 [stay.put-not-]-think.that-APPOS-3PL.OBJ
 ‘thinking they were not staying put.’

In traditional terms, the expression is a single Verb Word 2-16. But it also includes two units with conventionalized coherence and meaning: (a) the position 2

Verb Base *qacingqa*- ‘to stay put’, formed semi-idiosyncratically from a root *qacig*- ‘be easy, comfortable’ plus *-ngqa*- ‘be in a state of V’; and (b) the VV Postbase *-yuke*- (which becomes *-cuk*- here by morphophonological rules) ‘to think that V’, formed semi-idiosyncratically from two otherwise independently-attested VV Postbases, *-yug*- ‘to want or tend to’ and *-ke*- ‘to consider as V’. By adding Dixon’s and Aikhenvald’s criterion, we neatly characterize the holophrasis of the expression.

Our criteria in §5, particularly subspan repetition (see §5.6), did get at some of this, by noting that *-yuke*- ‘think that V’ could, even as a position 3 VV Postbase, select a span 2-7 out of turn. But if we were considering the corresponding English gloss, we would have had no problem calling ‘think(ing)’ a verb core. So why can’t we call *-yuke*- a verb core, especially since doing so might not only comport with Dixon and Aikhenvald’s criterion, but also unlock further useful criteria?

Recall that Tallman (2021: 13) defines the verb core “as a verb root *or* as a verb stem which would no longer remain of the same category if any of its affixes were stripped of”. He goes on to say, “The verb core constitutes the semantic head of the sentence insofar as the sentence is an example of a verbal predicate construction (see Croft 2001: 259; Anderson 2006: 211–27 on the concept of semantic head).” Meanwhile Croft (2001), recasting in semantic terms an already-present tension in the morphosyntactic notion of headedness (cf. e.g., Zwicky 1985), distinguishes notions of headedness based on lexical density (‘primary information-bearing unit’ (PIBU) in a constituent (p. 244) and PROFILE EQUIVALENT, an element within a complex expression that “profiles/describes a kind of the thing profiled/described” by the whole expression (p. 257). And Croft (2001) then defines the head as “the profile equivalent that is the primary information-bearing unit, that is, the contentful item that most closely profiles the same kind of thing that the whole constituent profiles” (p. 259). But then, when considering headedness ‘in morphology’, he stipulates (without argument) that within a word, “both inflection and root are profile equivalents of the whole; but the root is the PIBU” (p. 268) and that “profile equivalence is not helpful in defining morphological structure, in particular the root–affix structure of words.” (p. 269)

But if we are trying to gauge wordhood, we cannot presuppose it. We cannot require the verb core to be what the grammatical tradition has, in advance, stipulated as being a single root or a stem, however well-founded that stipulation may be. Nor can we say by fiat that the PIBU works one way within words and another way across words. Rather, we should allow elements other than roots or stems to be designated as verb core for purposes of constituency measurement, especially if they show headedness in either established sense: as a (relatively) lexically

dense PIBU within a given constituent; or as a profile-equivalent that may be seen as selecting or having scope over or determining a complement constituent (see Haspelmath 1992, in particular, for a defense and interesting synthesis of the long-established tradition of gauging headedness in morphology; also Woodbury 1981 for analytic program in terms of multiple or conflicting headedness notions). It even means rejecting, a priori, the notion of *primary* information-bearing unit itself, since it stacks the deck against holophrasis by presupposing a lack of multiple and perhaps equally primary information-bearing units within a given constituent. Rather, it may be better to consider or try to measure lexical density in a more general and abstract way.

So in the case of *-yuke-* 'think that' in (52), *-yuke-* evidently has some lexical density; and it also is the profile equivalent of the whole Verb Base to which it belongs, *qacingqanricuke-* 'to think O was not staying put'; for example, *-yuke-* is responsible for the transitivity of the whole expression. *-yuke-* is therefore a head in every sense.

Mechanically, allowing *-yuke-* in (52) to count as verb core (position 2) in our planar structure would require no re-working at all of the planar structure itself; as shown in (53), the part before it, *qacingqanrite-* 'to not stay put', would be relegated to the position 1 peripheral zone, and the remaining Inflection would conform to the planar structure as positions 12 and 15. It would also not preclude a planar level at which *-yuke-* counts as occupant of position 3, as originally done in (52).

(53)	<i>qacingqa -nri -cuk -lu -ki</i>
1{v}:	[2- 7] -3 -12 -16
2{v}:	[1 -] -2 -12 -16
	stay.put-not-think.that-APPOS-3PL.OBJ
	'thinking they were not staying put.'

But what it would do is present a class of cases where the span 2-16 sometimes fails to pass the constituency diagnostics that it passed in §5. That is, whenever a VV Postbase is reckoned as verb core, it will be the span 1-16 (and the span 1-20) that passes the constituency diagnostics, rather than 2-16 (and 2-20) as shown in Table 2.

Furthermore, if we consider certain VV Postbases as verb cores, we encounter many situations where their syntactic selection properties and semantic scope extends *beyond* just the Verb Bases they immediately follow. Consider:

- (54) *Uyurama tengmiaq tan'gurrarmun ivar-cit-a-a*
 1 1 1 2-3-12-16
 my.brother.REL.SG bird.ABS.SG boy.DAT.SG seek-let-IND-3SGA+3SG.OBJ
 'My brother made/let the boy look for the bird.' (Woodbury 1985: 274)

Here, the VV Postbase *-cit-* 'make/let' is a dense(ish) profile-equivalent for the whole phrase that is shown because—recalling Croft's definition—"it profiles/describes a kind of the thing profiled/described by the whole expression": it, by itself, specifies transitivity and the overall argument structure, introducing both the Relative case A that does the 'letting', and a complement proposition that expresses the 'seeking'. As such, the syntactic selection and semantic scope of *-cit-* extends beyond the Verb Base *ivar-* 'seek', because if we consider its own argument structure, then 1b-to-2 is a subspan over which the head *-cit-* 'let' in position 3 has scope. That is, we can distinguish a subspan from 1b-to-2, where all internal arguments are explicit.

Moreover, there also are NV Postbases—deriving a Verb Base from a Nominal Base—with lexically dense verbal meanings like 'have,' 'be,' 'wear,' 'hunt,' 'eat,' 'be tired of (N)', and others (but see discussion in Mithun 1998, who points out that for some NV Postbases, there are etymologically-unrelated Verb Bases with similar senses, but that are inherently more foregrounded in the discourse than the Postbase; relatedly, Johns 2007 argues such NV Postbases are light verbs). They too may be worthy of consideration as verb core, and as such lead to an extension of syntactic selection and semantic scope beyond the Verb Word that they head. For example:

- (55) *ciku-meng atauci-meng ene-ngqer-tu-a*
 1a-1b 1c-1d 1e-2-12-16
 ice-INS.SG one-INS.SG house-have-IND-3SG.S
 'I have one house made of ice.' (Woodbury 2017: 352)

In (55) we have taken the NV Postbase *-ngqerr-* 'to have N' to be the verb core (position 2): it shows a certain level of lexical density, as well as profile-equivalence not only for the Verb Base *enengqerr-* 'to have a house' but also for the whole phrase meaning 'have one house made of ice'. That is, *cikumeng* '(with) ice' and *ataucimeng* '(with) one' appear to be "stranded" as apparently separate words, shunted into the (accusative-functioning) Instrumental case. And yet, as semantic modifiers of the nominal head *ene-* 'house', they must come within the scope of *-ngqer(r)-* 'have'. At the same time the Verb Word *enengqertua* 'I have a house' conforms to all the tests that have been noted for positions 2-16 in §5. This

bracketing paradox has been discussed and debated extensively by Sadock (1980, 1991), Mithun (1984), Baker (1988), and many others, under the heading 'noun incorporation.' One of the issues under debate is the extent to which suffixes can count as verbs; and another is whether extended NP complements for such 'noun-incorporating' verbs—whether suffixal or not—should count as constituents even when they have one subconstituent (usually the head) inside a holophrastic word, and the remaining subconstituent(s) outside it, as in (54) and (55).¹⁵ Since our focus is constituency whether above, below, or across the putative word-level, it is important for us not to join such debates on a priori grounds, but rather to add to our battery of constituency diagnostics ones that can measure the validity of constituents without regard to putative wordhood.

In summary, we have pointed to three areas or dimensions of consideration where our program can be adjusted so that it better detects the constituency implications of holophrasis in YI languages and perhaps others. We can label and formulate them as follows:

LEXEMIC THRESHOLD. Calling on Dixon and Aikhenvald's criterion of 'conventionalized coherence and meaning', at what point do you 'not bother' with less-than-productive patterning when determining what elements (and hence positions) you will consider in formulating and applying a planar structure for the purpose of measurement? In the present analysis, for example, 'Bases' and 'Postbases' were the 'elements' and their internal composition was ignored; but would actually analyzing their component morphemes and assigning them to positions end up making Cup'ik Verb Bases and VV Postbases look more word-like? And are there languages for which such a strategy makes sense?

¹⁵Baker (1988) frames this as a head-to-head syntactic movement transformation, where (roughly) the lexical head (whether a word or an affix) of a subcategorized phrase subjoins to or incorporates with the head of the phrase that subcategorizes it. Sadock 1991's Incorporation Principle is the same idea, but treats the pre-incorporation constituency (heads-apart) as constrained by a Syntax module and the post-incorporation constituency (heads-together) as constrained by a Morphology module, motivating a constituency clash or bracketing paradox; see also Woodbury 1996 for an alternative formulation along similar lines. Meanwhile in the literature on Canadian Inuit varieties, Compton & Pittman (2010) go so far as to see Bases and Postbases as Words and the (traditional, span 2-16) Word as phrases with high degrees of phonological cohesion. Yuan (2018) follows Johns (2007) in seeing NV and VV Postbases as light verbs that 'get together' syntactically (or just postsyntactically) with their complements (or the heads of their complements) because they are in some sense light or relatively lexically un-dense; but Yuan also very importantly observes that not all NV Postbases select and combine with bare stems: although the fact are quite diverse across the YI languages, some NV Postbases combine with phrasal nominals (including personal pronouns and demonstratives) and some with oblique case marked words or phrases: See also Woodbury (1996) for discussion of some similar phenomena for Cup'ik.

PROFILE EQUIVALENCE. Calling on this aspect of headedness as dissected by Croft, would it be useful to re-formulate and then re-fit an alternative planar structure in which the ‘verb core’ is determined not merely by what is a Base, but instead in terms of strong profile equivalence within a larger domain? This may end up making Cup’ik Verb Bases and VV Postbases look more word-like, and likewise, reveal constituency patterns that include pieces of traditional Words together with other Words external to them, as discussed in the incorporation literature.

LEXICAL DENSITY: And calling explicitly on this other aspect of Croft’s dissection of headedness, can we perhaps motivate—in some cases at least—Tallman’s and Croft’s initial intuition that come what may, the verb core of a holophrastic word will contain a root rather than only an affix or affixes? Such a notion of lexical density might be reckoned relative to its contribution to the whole clause or phrase in which it occurs (Croft’s PIBU); or relative to sense relationships within a lexicon as a whole, such as relations of hyponymy or extensional inclusiveness; or even relative to phrasal pragmatic prominence, as Mithun (1998) has intriguingly proposed for Central Alaskan Yupik NV and VV Postbases, arguing that they have less pragmatic saliency than comparable Verb Bases that might paraphrase them.

Acknowledgments

I wish to thank the late Joe Friday, Leo Moses, Mary Moses, and Ulric Nayamin; and Rebecca Nayamin, John Pingayak, and many others in Chevak who have taught me what I know of Cup’ik. I also wish to thank Jerry Sadock, Willem de Reuse, Gladys Camacho Ríos, Richard Compton, Patience Epps, Nick Evans, Marianne Mithun, Osahito Miyaoka, Adam Tallman, Hiroto Uchihara and Natalie Weber for discussion of many of the issues raised here, and I am especially grateful to Adam for his thoughtfulness and skill in creating the program within which we are working in this volume. I gratefully acknowledge support for my work in Chevak from the National Science Foundation (grants SBR 9511856, BNS 8618271, and BNS 8217785).

Abbreviations

1	first person	A	agent
2	second person	ABS	absolute
3	third person	APPOS	apposition

CNTG	contingent mood	PL	plural
CONSEQ	consequential mood	POSS	possessive
CONTIN	continuative	QUOT	quotative
CONTMP	contemporative mood	REFL	reflexive
DAT	dative	REL	relative case
DU	dual	S	argument of
IND	indicative		intransitive verb
INS	instrumental	SG	singular
LOC	locative	TRPRT	transitive participle
OBJ	object		mood
OPT	optative		

References

- Anderson, Gregory D. S. 2006. *Auxiliary verb constructions*. Oxford: Oxford University Press.
- Baker, Mark. 1988. *Incorporation: A theory of grammatical function changing*. Chicago: University of Chicago Press.
- Bergsland, Knut. 1955. *A grammatical outline of the Eskimo language of West Greenland*. Zug: Micro-Editions of Interdocumentation.
- Boas, Franz. 1911. Introduction. In *Handbook of American Indian Languages*, vol. 1 (Bulletin 40), 1–10. Washington: Bureau of American Ethnography.
- Compton, Richard & Christine Pittman. 2010. Word-formation by phase in Inuit. *Lingua* 120. 2167–2192.
- Croft, William. 2001. *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- de Reuse, Willem J. 1994. *Siberian Yupik Eskimo: The language and its contacts with Chukchi* (Studies in Indigenous Languages of the Americas). Salt Lake City: University of Utah Press.
- de Reuse, Willem J. 2009. Polysynthesis as a typological feature: An attempt at a characterization from Eskimo and Athabaskan perspectives. In Marc-Antoine Mahieu & Nicole Tersis (eds.), *Variations on polysynthesis: The Eskaleut languages* (Typological studies in language 86), 19–34. Amsterdam; Philadelphia: John Benjamins.
- Dixon, R. M. W. & Alexandra Y. Aikhenvald. 2002. Word: A typological framework. In R. M. W. Dixon & Alexandra Y. Aikhenvald (eds.), *Word: A Cross-linguistic Typology*, 1–41. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511486241.002.

- Duponceau, Peter S. 1819. *Report of the Corresponding Secretary, read 12th January, 1819* (Transactions of the Historical and Literary Committee of the American Philosophical Society I). Philadelphia: A. Small.
- Foley, William A. & Robert D. Van Valin. 1984. *Functional syntax and Universal Grammar*. Cambridge: Cambridge University Press.
- Fortescue, Michael D. 1980. Affix ordering in West Greenlandic derivational processes. *International Journal of American Linguistics* 46. 259–278.
- Fortescue, Michael D. 1984. *West Greenlandic* (Croom Helm Descriptive Grammars). Kent: Croom Helm.
- Fortescue, Michael D. 1994. Polysynthetic morphology. In R. E. Asher (ed.), *The Encyclopedia of language and linguistics*, 2600–2602. Oxford: Pergamon.
- Fortescue, Michael D. 2015. *The sources of Eskimo-Aleut verbalizing affixes*. Unpublished manuscript.
- Fortescue, Michael D., Steven Jacobson & Lawrence Kaplan. 2010. *Comparative Eskimo dictionary, with Aleut cognates*. 2nd edn. Fairbanks: Alaska Native Language Center.
- Haspelmath, Martin. 1992. Gramaticization theory and heads in morphology. In Mark Aronoff (ed.), *Morphology now*, 69–82. Albany: State University of New York Press.
- Haspelmath, Martin. 2011. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica* 45(1). 31–80. DOI: 10.1515/flin.2011.002.
- Jacobson, Steven A. 1995. *A practical grammar of the Central Alaskan Yup'ik Eskimo language*. Fairbanks: Alaska Native Language Center, University of Alaska.
- Jacobson, Steven A. 2012. *Yup'ik Eskimo dictionary*. Fairbanks: Alaska Native Language Center, University of Alaska.
- Johns, Alana. 2007. Restricting noun incorporation: Root movement. *Natural Language and Linguistic Theory* 25(3). 535–576.
- Kiparsky, Paul. 1985. Some consequences of lexical phonology. *Phonology Yearbook* 2. 85–138.
- Kleinschmidt, Samuel. 1851. *Grammatik der Grönländischen Sprache*. Reprinted 1968, Hildesheim: Georg Olms. Berlin: Geog Olms.
- Lieber, Francis. 1853. The plan of thought in American languages. In Henry Rowe Schoolcraft (ed.), *History and statistics of the Indian tribes of the United States*, vol. 2, 346–349. Washington.
- Mattissen, Johanna. 2004. A structural typology of polysynthesis. *Word* 55. 189–216.

- Mattissen, Johanna. 2006. The ontology and diachrony of polysynthesis. In Dieter Wunderlich (ed.), *Advances in the theory of the lexicon*, 287–353. Berlin: De Gruyter.
- Mattissen, Johanna. 2017. Sub-types of polysynthesis. In Nicholas Evans, Michael D. Fortescue & Marianne Mithun (eds.), *The Oxford Handbook of Polysynthesis*, 70–98. Oxford: Oxford University Press.
- Mithun, Marianne. 1984. The evolution of noun incorporation. *Language* 60. 847–93.
- Mithun, Marianne. 1998. Yup'ik roots and affixes. In Osahito Miyaoka & Fubito Endo (eds.), *Languages of the North Pacific Rim*, vol. 4, 63–76. Kyoto: Kyoto University.
- Mithun, Marianne. 2009. Polysynthesis in the Arctic. In Marc-Antoine Mahieu & Nicole Tersis (eds.), *Variations on polysynthesis: The Eskaleut languages* (Typological studies in language 86), 3–18. Amsterdam; Philadelphia: John Benjamins.
- Miyaoka, Osahito. 1985. Accentuation in Central Alaskan Yupik. In Michael Krauss (ed.), *Yupik Eskimo prosodic systems: Descriptive and comparative studies* (Alaska Native Language Center Research Papers 7), 51–76. Fairbanks: Alaska Native Language Center, University of Alaska.
- Miyaoka, Osahito. 2012. *A grammar of Central Alaskan Yupik (CAY)* (Mouton Grammar Library 58). Berlin: Walter de Gruyter.
- Nichols, Johanna. 1986. Head marking and dependent marking grammar. *Language* 62. 56–119.
- Reed, Irene, Osahito Miyaoka, Steven A. Jacobson, Paschal Afcan & Michael E. Krauss. 1977. *Yup'ik Eskimo grammar*. Fairbanks: Alaska Native Language Center, University of Alaska.
- Sadock, Jerrold M. 1980. Noun incorporation in Greenlandic: A case study of syntactic word formation. *Language* 56. 300–319.
- Sadock, Jerrold M. 1991. *Autolexical syntax: A theory of parallel grammatical representations*. Chicago: University of Chicago Press.
- Sadock, Jerrold M. 2003. *A grammar of Kalaallisut (West Greenlandic Inuttut)* (Languages of the World/Materials 162). Munich: Lincom Europa.
- Sadock, Jerrold M. 2017. The subjectivity of the notion of polysynthesis. In Nicholas Evans, Michael D. Fortescue & Marianne Mithun (eds.), *The Oxford handbook of polysynthesis*, 99–114. Oxford: Oxford University Press.
- Sapir, Edward. 1921. *Language*. New York: Harcourt, Brace & World.
- Tallman, Adam J. R. 2020. Beyond grammatical and phonological words. *Language and Linguistics Compass* 14(2). 1–14. DOI: 10.1111/lnc3.12364.

- Tallman, Adam J. R. 2021. Constituency and coincidence in Chácobo (Pano). *Studies in Language* 45(2). 321–383. DOI: 10.1075/sl.19025.tal. (20 April, 2023).
- Woodbury, Anthony C. 1981. *Study of the Chevak dialect of Central Yup'ik Eskimo*. Berkeley: University of California at Berkeley. (Doctoral Dissertation).
- Woodbury, Anthony C. 1984. Eskimo and Aleut languages. In David Damas & William C. Sturtevant (eds.), *Arctic*, vol. 5 (Handbook of American Indians), 49–63. Washington: Smithsonian Institution Press.
- Woodbury, Anthony C. 1985. Marginal agent clauses in Central Alaskan Yupik Eskimo internal and external syntax. In William H. Eilfort (ed.), *Papers from the parasession on causatives and agentivity*, vol. 21, part 2 (Proceedings of the Chicago Linguistic Society), 271–292. Chicago Linguistic Society.
- Woodbury, Anthony C. 1987. Meaningful phonological processes: A consideration of Central Alaskan Yupik Eskimo prosody. *Language* 63(4). 685–740.
- Woodbury, Anthony C. 1996. On restricting the role of morphology in Autolexical Syntax. In Eric Schiller, Barbara Need & Elisa Steinberg (eds.), *Autolexical syntax: Ideas and methods*, 319–363. Berlin: Mouton.
- Woodbury, Anthony C. 2002. The word in Cup'ik. In R. M. W. Dixon & Alexandra Y. Aikhenvald (eds.), *Word: A cross-linguistic typology*, 79–99. Cambridge: Cambridge University Press.
- Woodbury, Anthony C. 2005. Morphological orthodoxy in Yupik-Inuit. In Marc Ettlinger, Nicholas Fleisher & Mischa Park-Doob (eds.), *Proceedings of the Thirtieth Annual Meeting of the Berkeley Linguistics Society*. Berkeley: Berkeley Linguistics Society.
- Woodbury, Anthony C. 2017. Central Alaskan Yupik (Eskimo-Aleut): A sketch of morphologically orthodox polysynthesis. In Nicholas Evans, Michael D. Fortescue & Marianne Mithun (eds.), *The Oxford handbook of polysynthesis*, 536–559. Oxford: Oxford University Press.
- Woodbury, Anthony C. & Jerrold M. Sadock. 1986. Affixial verbs in syntax: A reply to Grimshaw and Mester. *Natural Language and Linguistic Theory* 4. 229–244.
- Yuan, Michelle. 2018. M-merger and copy spell-out in Inuktitut noun incorporation. In *Proceedings of NELS* 48.
- Zwicky, Arnold M. 1985. Heads. *Journal of Linguistics* 21. 1–29.