
**Constraints on compounds and incorporation**

Marianne Mithun  
University of California, Santa Barbara

This chapter examines the status of a type of compounding often cited as a construction that straddles the boundary between morphology and syntax. Noun incorporation constructions in Kapampangan, Mohawk, and Central Alaskan Yup'ik Eskimo are evaluated with respect to the No Phrase Constraint and the Lexical Integrity Hypothesis. Though they share certain fundamental properties, the constructions differ in crucial ways, reflecting differing stages of diachronic development.

1. **Introduction**

A central issue within morphological theory has been the delimitation of the domain and its relation to syntax. Two hypotheses have been central to the discussion, traced in detail in Lieber and Scalise 2006. The No Phrase Constraint, described by Botha (1981: 18), denies morphological operations access to the output of syntax: words cannot be formed from syntactic phrases. The Lexical Integrity Hypothesis, first proposed by Lapointe (1981: 22), denies syntactic operations access to the internal structure of words. Compounding has played a prominent role in these discussions, in part because of its position at the border between the two domains. On one side, it produces lexical items, a classic morphological function. On the other, it can take words as its input, and, in some languages, phrases which are the products of syntactic operations.

---

* All examples are cited in the practical orthographies in use in the communities, but the phonetic values of the symbols are generally not far from standard IPA conventions. Kapampangan writers usually spell clitics as separate words. The digraph <ng> represents the velar nasal <ŋ>. For Mohawk, the letter <i> represents a palatal glide before a vowel, and a high front unrounded vowel otherwise. The digraphs <en> and <on> represent nasalized vowels, caret and high back rounded u respectively. The colon <ː> indicates vowel length, the acute accent ⟨ā⟩ stress with high or rising tone, and the grave accent ⟨à⟩ stress with falling tone. The apostrophe ⟨’⟩ represents glottal stop. For Yup’ik, the symbols <g> and <r> represent velar and uvular fricatives respectively, and the digraphs <gg> and <rr> their voiceless counterparts, except before a voiceless consonant where they automatically assimilate the voicelessness. The letter <e> represents a schwa.
One type of compounding that seems particularly syntactic is noun incorporation, the combination of a noun and a verb to form a larger verb. A typical example is the Mohawk verb stem -\textit{iti-ienta} - 'fish-get' = 'to catch fish.' In work on the morphology-syntax interface, Li has proposed that parts of words 'are invisible to syntax unless there is a thematic relation expressed sublexically' (2005: 4). If any kind of compounding constitutes such a case, noun incorporation should be a prime candidate, since it consists of a verb and a noun that bears some relation to it.

Here it is shown that noun incorporation can be characterized by certain basic properties, but it is not homogeneous cross-linguistically, nor is it necessarily static through time. Examples are drawn from three genetically and areally unrelated languages: (1) Kapampangan, an Austronesian language of the Philippines; (2) Mohawk, an Iroquoian language of northeastern North American, and (3) Central Alaskan Yup'ik, an Eskimo-Aleut language of Alaska. The constructions are first described, then evaluated with respect to the No Phrase Constraint (the prohibition against syntactically complex non-heads, here incorporated nouns) and the Lexical Integrity Hypothesis, specifically the claim that parts of words (here incorporated nouns), cannot serve as antecedents of pronouns). Differences among the constructions in the three languages are related to properties associated with different stages of diachronic development.

2. The constructions

Incorporating constructions in the three languages share certain properties. All consist of a noun and a verb which constitute a larger verb or verb phrase. Each is formally intransitive. The fish are not syntactic arguments, nor are they distinguished for number, definiteness, or case.

\begin{enumerate}
\item 'They caught fish.'
\begin{tabular}{ll}
Kapampangan & \textit{Ikualang asan}. \\
Mohawk & \textit{Wahatit siai:ta'ne}. \\
Yup'ik & \textit{Neqetut}.
\end{tabular}
\end{enumerate}

In what follows, all examples attributed to speakers are from spontaneous speech.

2.1 Kapampangan

Argument structure is indicated in Kapampangan in three ways: (i) by the verbal morphology, (ii) by case-marked pronominal enclitics, and (iii) by case-marked determiners. Verbs are inflected for transitivity, voice, and tense/aspect/modality with combinations of prefixes, infixes, and suffixes. The verb \textit{labas} 'pass by' has the intransitive (\textit{INTR}) form \textit{lumabas} in (2a), but the transitive (\textit{TR}) form \textit{labas napan} in (2b). Pronominal enclitics identify the core arguments of the clause, whether or not they are further identified by
lexical nominals. Some of the clitics are fused forms. The clitic \textit{ku} in the intransitive clause ‘I’ll pass by’ refers to just one argument: absolutive ‘I’. The clitic \textit{ke} (\textit{ku}+\textit{ya}) in ‘I’ll pass their house’ refers to two: ergative ‘I’ and absolutive ‘it’ (the house). The determiner \textit{ing} before ‘their house’ identifies this constituent as the absolutive argument.

(2) Kapampangan clause structure: Bernadette Mangasar, speaker p.c.

a. \textit{Lumabas}=\textit{ku}.
   \hspace{1cm} \text{pass.INTR}=1\textit{SG.ABS}
   \hspace{1cm} ‘I’ll pass by.’

b. \textit{Labasnan}=\textit{ke} \hspace{1cm} \textit{ing} \hspace{1cm} \textit{bale da}.
   \hspace{1cm} \text{pass.TR}=1\textit{SG.ENG}/3\textit{SG.ABS} \hspace{1cm} \text{ABS.SG.LK} \hspace{1cm} \text{house} \hspace{1cm} 3\textit{PL.POSS}
   \hspace{1cm} ‘I’ll pass their house.’

The pronominal clitics identify only core or syntactic arguments: ergatives and absolutes. The determiners on noun phrases distinguish three cases: ergative, absolutive, and oblique. In addition to case, the determiners also mark singular/plural and common/proper distinctions.

(3) Kapampangan clause structure: Clemente Roman, speaker p.c.

\textit{Ing} \hspace{1cm} \text{Bataan, sinuk=ya} \hspace{1cm} \textit{karing} \hspace{1cm} \textit{Apones}.
\hspace{1cm} \text{ABS.SG} \hspace{1cm} \text{Bataan} \hspace{1cm} \text{surrendered.INTR}=3\textit{SG.ABS} \hspace{1cm} \text{OBL.PL.PROL.LK} \hspace{1cm} \text{Japanese}
\hspace{1cm} ‘Bataan surrendered to the Japanese.’

Constituents of a phrase may be linked by the enclitic \textit{=ng} (velar nasal \textit{=y}). In (4) the linker (\textit{Lk}) relates the constituents of ‘the old woman’.

(4) Kapampangan linker \textit{=ng}: Clemente Roman, speaker p.c.

\textit{Migtakaya} \hspace{1cm} \textit{ing} \hspace{1cm} \textit{matu mang} \hspace{1cm} \textit{babai}.
\hspace{1cm} \text{Migtaka}=\textit{ya} \hspace{1cm} \text{[}=\textit{ng} \hspace{1cm} \text{matua}=\textit{ng} \hspace{1cm} \text{babai}.]
\hspace{1cm} \text{surprised.INTR}=3\textit{SG.ABS} \hspace{1cm} \text{ABS.SG}=\textit{Lk} \hspace{1cm} \text{old}=\textit{Lk} \hspace{1cm} \text{woman}
\hspace{1cm} ‘[The old woman] was surprised.’

The structure of the Kapampangan ‘they caught fish’ in (1) can now be appreciated. The form of the verb \textit{ikua} ‘got’ is intransitive. The only argument of the clause is the absolutive \textit{=la} ‘they’. The fish have no syntactic status: they are not represented by a pronominal clitic, and there is no determiner before the noun. The noun is linked directly to the verb.

(5) Kapampangan incorporation: Clemente Roman, speaker p.c.

\textit{Ikualang} \hspace{1cm} \textit{asan}.
\hspace{1cm} \text{ikua}=\textit{la}=\textit{ng} \hspace{1cm} \text{asan}
\hspace{1cm} \text{got.INTR}=3\textit{PL.ABS}=\textit{Lk} \hspace{1cm} \text{fish}
\hspace{1cm} ‘They caught fish.’
2.2 Mohawk

Mohawk verbs consist minimally of a pronominal prefix and a verb root. The pronominal prefix identifies the arguments of the clause: one for intransitives and two for transitives (provided both are animate). The verb in (6) contains the pronominal prefix -hi- 'I/him'.

(6) Mohawk clause: Kahentoréthha Marie Cross, speaker p.c.
   *Wahtie:hte*
   *wa-hi-ie-htí-e*
   FACTUAL-LSG/MSG-awake-CAUS-PRF the
   *riién:a*
   *ni-ién=a*
   the LSG/MSG-have.as.offspring=DIM
   I awakened him
   'I woke my son up.'

Nouns are unmarked for number or case.

Verbs may also contain various other prefixes and suffixes, as well as an incorporated noun stem. The incorporated noun occurs after the pronominal prefix and immediately before the verb root. If the noun stem ends in a consonant, and the following verb root begins in a consonant, a linker vowel -a- is inserted between the two.

(7) Mohawk incorporation: Sonny Edwards, speaker p.c.
   *Wahtitsiaén:tañe*.
   wa-hati-itsi-a-ient-añ-e'
   FACTUAL-M.PL.AGT-fish-LK-have.INCHOATIVE-PRF
   'They caught fish'

2.3 Central Alaskan Yup'ik

All Yup'ik verbs begin with a root, potentially followed by one or more derivational suffixes. They end with a mood and pronominal suffixes. Most of the mood suffixes, including the indicative, distinguish transitivity. The pronominal suffix refers to the core or syntactic arguments of the clause: one for intransitives and two for transitives.

(8) Yup'ik clause: Elena Charles, speaker p.c
   *Tangerruryartullruaq*.
   tangerr-ur-yartur-llru-ar-qa
   see-purposely-go.in.order.to-PAST-TR.INDIC-1SG/3SG
   'I went to see him.'

Noun suffixes specify number, possessor, and case: ergative, absolutive, allative, ablative, locative, vialis ('through'), and aequalis ('like').

Some Yup'ik verbs are based on a noun root followed by a derivational verbalizing suffix with relatively concrete meaning, such as 'hunt', 'gather', 'eat', 'say', 'buy', 'smell or taste like', 'encounter', 'acquire', 'lack', 'hit', 'cook', 'make', and more (Mithun 1998a,b). The Yup'ik 'they caught fish' consists of an initial noun root neqe- 'fish, food' followed
by a verbalizing suffix -te- 'catch.' It is clear from both the intransitive mood suffix -u-
and the third person plural absolutive pronominal -t that the clause is grammatically
intransitive.

(9) Yup'ik
   Neqtut.
   neqc-te-u-t
   fish-catch-INTR.INDIC-3PL.ABS
   ‘They caught fish.’

3. Internal structure

The 'catch fish' constructions in all three languages are endocentric. All consist of a
nominal element and a verbal element. The verbal element functions as the head;
structurally, its lexical category is passed on to the construction as a whole; semanti-
cally, it is a hyponym of the whole: its meaning is modified or narrowed by the non-
head: catching fish is a kind of catching. The constructions might qualify as synthetic
compounds according to classical definitions: their nominal constituents seem to 'bear
a thematic role in relation to that verb stem identical or very similar to the role it has
in a corresponding verb phrase' (Carstairs-McCarthy 1992: 109) or 'in a sentence,
could function as an argument of that verb' (Bauer 2003: 44).

In all of the languages, the incorporating constructions have analytic counter-
parts. (The two structural alternatives are not always equally idiomatic with specific
noun-verb combinations, but the two structures co-exist). Often the noun in the ana-
lytic counterpart is a syntactic argument. In (10), for example, the fish is the absolutive
of a transitive.

(10) Kapampangan alternatives: Bernadette Mangaser, speaker p.c.
   Selakatne        ing      asan.
   selakat=na=ya i=ng      asan.
   trapped.TR=3SG.ERG=3SG.ABS 3SG.ABS=LK fish
   ‘He trapped the fish.’

Only definite nominals can function as syntactic arguments in Kapampangan. Indefi-
nites appear either in incorporating constructions or as obliques.

The roles of incorporated nouns are not limited, however, to those which could be
interpreted as syntactic arguments. They are, furthermore, semantically heteroge-
neous: often semantic patients or goals, but also instruments, locations, and others
more difficult to specify.
(11) Kapampangan
a. *Pota midinanlang sabun.*
   later they get splattered soap
   ‘They might be splattered with soap.’

b. *Mitambunanta nang basura.*
   we will be buried already garbage
   ‘We’ll just be buried in garbage.’

(12) Mohawk
a. *kahserei‘tāneren’*
   ‘It is string-tied’ = ‘it is tied up with string’

b. *onke‘nionhsökha’*
   ‘I nose-leak’ = ‘I have a runny nose’

c. *thiwa‘kate‘noniōni*
   ‘I’ve just knot-made myself’ = ‘I’m all hunched up’

(13) Yup’ik: Jacobson 1984
*iarpigkuartua*
‘I’ll sea-go.by.way.of = I’ll go by sea.’

It is well known that the semantic relationships between the constituents of noun-noun compounds are not generally predictable. The non-heads of English endocentric noun-noun compounds, for example, need only be ‘appropriately classificatory’ in the sense of Downing (1977). Relationships between members of synthetic compounds are generally thought to be more constrained: the verbal element ‘usually determines unequivocally the meaning of the compound, thus ruling out other readings’ (Scalise 1984: 90). The interpretation of the incorporating constructions described here is not specified by either the grammar or the verbal head. Incorporated nouns with different semantic roles can appear with the same verb root.

(14) Mohawk verb root -*itakahke-*
*ratti‘titākhkhe’*  ‘they were body-moving’ = ‘they were riding’

*ka‘nerohkwitākhkhe’*  ‘it is box-moving’ =
‘there’s a box inside carried along’

*tahonatha‘hitakahkhe’*  ‘they were road-moving hither’ =
‘they came walking down the road’

4. Transitivity and argument structure

The constructions seen so far have all been intransitive, and in some languages this is the only possible product of noun incorporation. In the three languages under discussion here, noun incorporation can produce both intransitive and transitive verbs.
The majority of Kapampangan verb roots have both intransitive and transitive forms, often multiple forms of each. The same is true of incorporating constructions. The sentence in (15) is transitive.

(15) Kapampangan transitive: Bernadette Mangaser, speaker p.c.
    Sekedakaming gebara.
    they put us barge
    ‘They put us on a barge.’

Some Mohawk verb roots are inherently intransitive, some transitive, and some are used both ways. The same is true of verb stems containing incorporated nouns. The transitivity is a lexical feature of the whole. The incorporated noun may or may not have any relation to a core argument of the clause. In (16), the incorporated -iar- ‘bag’ narrows the meaning of the verb ‘turn’ to actions appropriate for mattresses. The clause is transitive.

(16) Mohawk transitive: Kahentó:ráhtha’ Marie Cross, speaker p.c.
    Ensiararakhathóhseron kaiarahrónion.
    again I will bag-turn them bags sitting here and there
    ‘I was going to flip the mattresses.’

The verb in (17) contains an incorporated noun -na’tsi- ‘pot’, and the clause is transitive, but that noun has no relation to the clausal argument ‘white beans’. It is just an element of the stem -na’tsi-a-niiont- ‘pot-suspend’ = ‘cook (in an iron pot over a fire)’.

    Tanon’ kará:ken nikasahe’tó:ten ensewanda’tsianiion:ten’.
    and white so it is a kind of bean you will pot-hang
    ‘And you’ll cook the white beans (in a big, iron pot).’

The transitivity of the Yup’ik constructions is also a lexical matter. Some verb roots are used only intransitively, some only transitively, but most both ways. The constructions under discussion here show the same variation. A number of derived N-ir- stems exist: kave-ir- ‘hail’, puy-ir- ‘smoke-ir- ‘be smoky’, neg-ir- ‘snare-ir- ‘set a snare’, ken-ir- fire-ir- ‘build a fire under something’ or ‘cook something’, ac-ir- name-ir- ‘name something’ mur-ir- wood-ir- ‘stoke something’, and more (Jacobson 1984: 457–8). The semantic roles of the nouns vary. Some of the derived stems are intransitive, some transitive, and some either, as in (18).

    a. atsirtuq
       atsa-ir-tu-q
       berry-provided-INTR.INDIC-3SG
       ‘it is well provided with berries’
b. *atsiraa*
   atsa-ir-a-a
   berry-provide-tr.indic-3sg/3sg
   'she added berries to it

Because incorporation can alter the argument structure of a verb, in all three languages it can provide alternatives that speakers exploit for discourse purposes. The Kapampangan speaker cited in (19) could have said 'Remove the leaves from the trees', casting the leaves as a core argument and the trees as oblique. Instead, he said 'Strip (leaf-remove) the trees', casting the trees as the core argument. The trees, important to the discussion at hand, continued as a core argument of the following clause.

(19) Kapampangan: Clemente Roman, speaker p.c.
   Gisandalang   piglako   bulung
   Gisang-da-la-ng   piglako   bulung
   finished=3pl.erg=3pl.abs=lk removed leaf
   'They stripped the trees of their leaves
   at *balát di-ng taná:man.*
   and killed common.pl.abs=lk tree
   and killed them.'

Kapampangan constructions with this effect generally lack the linker =ng.

Incorporation in Mohawk provides speakers with similar options. Asked whether she knew a certain person, the speaker in (20) replied 'I've name-heard him' = 'I've heard of him' rather than 'I've heard his name.' The topic of conversation was the person, not his name.

(20) Mohawk topical person as argument: K. Jacobs, speaker p.c.
   Rihsennahrónkha'.
   ri-hsenn-ahronk-ha'
   m.sg/1sg-name-hear-habitual
   I used to name-hear him = 'I've heard of him.'

Similar alternations are available to Yup'ik speakers. Discussing a hunting trip, the speaker in (21) reported that a hunter 'wing-hit the bird' rather than 'he hit its wing.' The bird, cast as a core argument (absolutive), was of more interest to the hungry hunters at that point than the wing.

(21) Yup'ik topical argument: George Charles, speaker p.c.
   Yaqulek   yaquartaa
   yaqu-lek   yaquarte-a-a.
   wing-thing.with   wing-hit-tr.indic-3sg/3sg
   bird   he wing-hit it
   'He hit the bird right in the wing.'
In all three languages, both intransitive and transitive verbs can enter into the construction, and the incorporating constructions can be intransitive or transitive.

5. The no phrase constraint and syntactically complex constituents

The No Phrase Constraint states that morphological operations should have no access to the output of syntactic operations. If compounding is a morphological process, it should not be possible to form compounds from syntactic phrases. Yet compounds with syntactically complex constituents are well documented, such as the English pipe and slapper husband and God is dead theology (Botha 1985, Lieber 1992, Lieber and Scalise 2006). The constructions under discussion here vary with respect to this property.

In the Kapampangan construction, syntactically complex nominals are in fact quite common. A typical example is (22).

(22) Kapampangan syntactically complex nominal: C. Roman, speaker
Gawa-la-ng tahada-ng mani at letsi plan.
will.make=3PL.ABS=LK brittle=LK peanut and milk flan
'They’ll make peanut brittle and milk flan.'

In Mohawk, incorporated nouns are never complex syntactically. Some are complex morphologically, but these are already well-established lexical items. The verb in (23) contains the incorporated stem ‘table’. This stem is a nominalized verb with an incorporated noun of its own: ate-khw-a-hra-MIDDLE-food-LINKER-set.

(23) Mohawk complex nominal: Kaia’titáhkhe’ Jacobs, speaker p.c.
katekhwahra’-tsheria’ákhons
k-atekhwahra’-’tsher-ia’ák-hon-s
1SG.AGT-table-NOMINALIZER-hit-DISTRIBUTIVE-HABITUAL
'I’m table pounding’ = ‘I’m pounding on the table.’

Some Mohawk incorporation might appear to involve syntactically complex nominals as input. On the basis of the English translation, the example in (24) might be interpreted as the result of incorporating just the head of a phrase ‘other story’, leaving the remainder ‘stranded’ outside.

(24) Mohawk stranded modifier?: Watshenni’ne Sawyer, speaker p.c.
Shé:kon ə́ia’ wakká:raíen’
shēkon ohia’ wak-kar-a-ien-
still other 1SG.PAT-STORY-LK-have-STATIVE
still other 1 story have
‘I have another story.’
But ò:ià' 'other' also appears with verbs without an incorporated noun, such as 'wear' in (25). It is a referring expression on its own: 'other ones'.

(25) Mohawk independent nominal: Margaret Edwards, speaker p.c.
    Enwà:ton' ò:ià' entsitewatste'
    en-w-aton-' ohià' en-tsi-tewa-at-st-e'
    FUTURE-N.AGT-be,possible other FUTURE-REP-1.INCL.AGT-MIDDLE-use-PRF
it will be possible other we will wear
'Ve can change clothes.'

Quantifiers and numerals occur with verbs containing incorporated nouns.

(26) Mohawk quantifier è:so' 'much': Watshenní:ne Sawyer, speaker p.c.
    Iah è:so' teionkwahwistaien'.
    iah eso' te-ionkwa-hwist-a-ien-
    not much NEG-1PL.PAT-money-LK-have-STATIVE
    not much did we money have
'Ve didn't have much money.'

But the same quantifiers also occur without an incorporated noun.

(27) Mohawk quantifier è:so' 'much': Watshenní:ne Sawyer, speaker p.c.
    Iah è:so' tekaién:tahkwe'
    iah eso' te-ka-ient-ahkwe'
    not much NEG-NEUTER.AGT-lie-Former.PAST
'There wasn't much.'

It is possible in Mohawk to evoke a whole event or fact, normally expressed in a clause, within a verb, but this is accomplished by a simple incorporated noun root like -rihw- 'fact, matter, affair, idea, etc.'

(28) Mohawk incorporated stand-in for a clause: Lazarus Jacob, speaker
    lorìhwwanón:nhste' tsi ni:hati:iere'
    io-rihw-anon-hst-e' tsi ni-hati-iер-e'
    N.PAT-matter-confidential-CAUS-STATIVE so PRT-M.PL.AGT-do-STATIVE
it is matter-confidential so they did
    ne onhwentsakaìon:ne ne ratitsihénhstasi.
the old world place the priests
'What the priests did in Europe has been kept confidential.'

In Yup'ik, syntactic phrases do not serve as stems for verbal derivation. The nominal base can be complex morphologically.
(29) Yup'ik morphologically complex noun stem: E. Charles, speaker p.c.
Enecuarirluki.
ene-cuar-ir-lu-ki
house-small-build-SUBORD-3PL/3PL
'They would build them small houses.'

Yup'ik does exhibit constructions which might appear to be derived from syntactically
complex NPs by head movement, stranding a modifier.

(30) Yup'ik stranded modifiers? Elena Charles, speaker p.c.
Qantangqerralliriit
qanta-ngqerr-lar-lrii-t
bowl-have-customarily-INTR.PARTICIPIAL-3PL
they bowl-have
'They have large bowls.'

The word *angelrianek* 'large', however, is a nominal in its own right. It is a nominalized
verb, best translated 'large ones'. Similar nominalized forms appear in construction
with verbs containing no noun root at all.

(31) Yup'ik similar 'modifiers' without no noun base: E. Ali, speaker
Assitelriamek-gguq
assite-ria-mek=gguq
be.bad-NMZ-ABL.PL=HEARSAY
bad one
'They sold him a bad one.'

As in Mohawk, numerals can appear in sentences with verbs based on a noun stem. Here,
too, however, the same numerals also appear with verbs containing no noun stem.

(32) Yup'ik stranded numeral?: Elena Charles, speaker p.c.
Tutgarangqerrasaqwa,
tutar-ngqerr-yaaqe-u-a
grandchild-have-actually-INTR.INDIC-1SG
I actually grandchild-have
'I actually have two grandchildren in Seattle.'

(33) Yup'ik numeral without noun base: Elena Charles, speaker p.c.
Malrugnek
malrug-nek
two-ABL.DU
two
'Ve stayed two nights.'
6. The lexical integrity hypothesis and sublexical reference

Lieber and Scalise 2006 provide a useful survey of the literature on the Lexical Integrity Hypothesis. As characterized by Anderson (1992: 84) this says that "The syntax neither manipulates nor has access to the internal structure of words". More recently, Ackema and Neeleman (2005: 109) have proposed that "The structure of complex words is invisible to syntax because syntax builds up the 'host structure' and morphologically complex words are inserted into this structure". Booij (2009) argues, on the other hand, that morphological and syntactic structure are built up in parallel, so 'that the possibility of syntax and semantics having access to word-internal structure is to be expected'. An indication that syntactic processes have access to the internal structure of words would be sublexical reference, in our case, reference to incorporated nouns.

In some incorporating constructions in Kapampangan, the constituents are linked by the enclitic =ng. In others, they are simply juxtaposed. When the linker is present, the noun is not an argument, but subsequent pronominal reference to the entity it introduces is very common.

(34) Kapampangan linked nominal as antecedent: C. Roman, speaker

\[Kumalang \quad mitsa, \quad sindian=de.\]
\[kum=a=la=ng \quad mitsa, \quad sindian=de.\]
\[get=3PL.ABS=LK \quad wick \quad ignite=3PL/3SG\]

'They will get a wick and ignite it.'

In constructions without the linker, the noun is again not an argument, and there is no evidence that the incorporated nominal ever serves as an antecedent for subsequent reference. When speakers wish to refer to an entity evoked in this construction, the noun is repeated.

(35) Kapampangan repetition of noun: Clemente Roman, speaker p.c.

['They will start tinkering with the pig's ears.]

\[Paqlakoro \quad dinat.\]
\[Paqlako=ro \quad dinat\]
\[remove=3PL/3PL \quad dirt\]
\[they will \quad dirt\text{-}remove \quad them\]

\[Ampo itang pambuk na kini,\]
\[and \quad then \quad its \quad snout,\]

\[paqlakore \quad dinat.\]
\[paqlako=re \quad dinat\]
\[remove=3PL/3SG \quad dirt\]
\[they \quad will \quad dirt\text{-}remove \quad it\]

'They'll clean the dirt off of its ears and clean the dirt off its snout.'
In Mohawk, incorporated nouns also do not establish reference for subsequent pronominal reference in spontaneous speech. To establish an antecedent for subsequent reference, an independent nominal is supplied.

(36) Mohawk subsequent reference, lexical nominal: S. Edwards speaker
Shaià:ta        ronkwe’tarakwen
one male person  he person-chose him
‘He picked one man

ne nen ówera’ enhatén’nikón:raren.’
the that wind  he will his mind-set on it
who would guard the wind.’

Incorporated nouns are not themselves referential, though they may evoke an identifiable referent, like ‘mind’ in the verb ‘guard’ (‘mind-set’) in (36). If a given incorporated noun occurs repeatedly within a short span of discourse, each instance does not necessarily evoke the same referent. In (37), speakers were discussing a book they were translating. The same book was being translated into other languages, but the Mohawk version was scheduled to be published first. The root -wenn- is not coreferential.

CB  Ohén:ton  î:kate’  onkwawén:na’.
in front  it is standing  our language
‘Our language is the first in line.’

JH  [...]  ni:  ónhka’  entionkhiatéwennátahe’  [...]  
also who  one will put one’s voice in to us hither
‘[Yes, it will be for sale], and whoever calls us on the telephone,
[we’ll have it in stock for them to purchase it.]’

CB  [...]  ahatíwennahnotómnion’
they could word stand variously
‘[That’s right, isn’t it, and they can really] read it.’

The noun bases in the Yup’ik constructions are also not usually picked up referentially by pronominals. In order to establish a referent, speakers usually supply an independent noun, often the same as the base.

(38) Yup’ik lexical nominal for subsequent reference: E. Charles, speaker
Iciugg’,  Frankiq  angyangqeltrul’.
remember Frank  he  boat-had
‘Remember Frank had a boat.

Angyaq-ilu  camek  imainani.
boat-and  things it contents-lacked
And the boat was empty.’
Yup'ik constructions like that in (38), with light verbs like 'have', 'exist', 'arrive', 'make', etc., are often used to introduce the idea of an entity into the discourse, normally followed by establishment of the referent in an independent lexical nominal. On relatively rare occasions, speakers skip the nominal.

(39) Yup'ik rare antecedent in light presentative: E. Charles, speaker p.c.

\(\text{Ucilirluteng, avegluku.}\)

they load-made they halved it

'They made a load, dividing it in half.'

The constructions in the three languages thus differ in the possibility of sublexical reference. Nominals in the Kapampangan =ng construction can and often do serve as antecedents for subsequent pronominal reference. Those in which the nominals are simply juxtaposed to verbs do not. Mohawk incorporated nouns do not normally establish reference. Yup'ik noun bases rarely establish reference.

7. Wordhood

The incorporating constructions in the three languages thus differ with respect to both the No Phrase Constraint and the Lexical Integrity Constraint. The variation is not random. A major function of morphology is word formation. The processes described here vary in the degree to which they create words.

The Kapampangan construction consists of the juxtaposition of full words, each recognizable and pronounceable in isolation. There are no phonological processes that apply specifically to these constructions, such as a compound stress rule. Speakers may pause between the verbal and nominal constituents. In the transcription of (40), three dots [...] represent a substantial pause, and two dots [...] a shorter pause. The verb and noun in the second line were separated by a substantial pause, and those in the third line by another, briefer pause.

(40) Kapampangan pauses: Clemente Roman, speaker p.c.

\(\text{Magdalalang sabun},\)

'They will bring soap.

\(\text{magdalalang} \quad \ldots \quad \text{pintura},\)

they will bring \ldots paint,

\(\text{magdalalang} \quad \ldots \quad \text{miyaliwang gamit}\)

they will bring \ldots other things [to clean the graves of their relatives].'

The verbal and nominal constituents in such constructions are sometimes even separated by additional words, such as the particle \textit{pa} 'also' in (41).
(41) Kapampangan intervening word: Clemente Roman, speaker p.c.

At biasaya pa-ng manese.
and she knew also-TK cook.rice
'And she also knew how to cook rice.'

Mohawk noun incorporation is a process of stem compounding. Neither the incorporated noun stem nor the incorporating verb stem can stand alone as a word, and speakers would not normally recognize either in isolation. Independent nouns generally contain a prefix indicating gender or a possessor, and a noun suffix. The noun root for 'fish' is -itsi-, but the noun word for 'fish' is kénsion'. Verbs must contain, minimally, a pronominal prefix identifying their core arguments, as in 'they caught it'.

(42) Mohawk noun

kénsion' wahatiién:ta'ne'.
ka-itsi-o' wa-hati-ienta'n-e'
neuter-fish-noun.suffix factual-pl.agt-catch-prf
'fish' 'they caught it'

There is no question that verbs with incorporated nouns, like wahatitsiaiën:ta'ne' 'they caught fish', are single words both phonologically and morphologically. Stress is consistently penultimate (with certain epenthetic vowels not counted, including the linker -a-). Numerous phonological processes occur word-externally but not between words. Within words, apart from the insertion of the linker -a-, the same kinds of phonological processes operate between incorporated noun stems and incorporating verb stems as elsewhere. Morphologically, both nouns and verbs show strict templatic ordering of morphemes. Asked to repeat sentences word-by-word, speakers always pronounce verbs with incorporated nouns as single units. They never pause inside of words, and words are never inserted inside of words.

The Yup'iK constructions are also clearly single phonological and morphological words. The word constitutes a phonological domain for stress and numerous other word-internal saadhi phenomena. Words also have strict word-internal morphological structure. All verbs begin with a root. Speakers never pause within words then continue, and they never insert other words inside of them.

8. Frequency and productivity

A difference often observed between morphological and syntactic processes is productivity. Degree of productivity is sometimes assessed in terms of frequency: highly productive processes are assumed to result in greater frequency of occurrence. The constructions in the three languages do vary in their frequency, though perhaps not in the way that might be expected. The figures in (43) contrast the densities of incorporating constructions in comparable 3000-word samples of spontaneous speech in the three
languages, primarily conversation. Incorporation is significantly more frequent in Mohawk than in the other two languages.

(43) Frequency: percentages of clauses with the construction
Kapampangan 7.5%
Mohawk 38.3%
Yup’ik 7.3%

But comparisons of actual productivity reveal a different pattern.

The Kapampangan construction shows full productivity: any noun phrase can be combined with any verb so long as it makes sense. Loanwords are easily and spontaneously accommodated.

(44) Kapampangan loan: Bernadette Mangaser
*Ibaltumeng aluminum foil.
‘Wrap it in aluminum foil.’

Mohawk differs substantially. Noun incorporation is productive, as can be seen in (45), with the incorporated loan job plus nominalizer -tsher.

(45) Mohawk incorporated loan
*Ronvatijobtsherawifne.
they had job-given them
‘They had given them jobs.’

But innovation is relatively rare. Speakers generally notice neologisms and comment on them. This particular combination ‘job-give’ has now found a place in the language as a lexical item.

The productivity of Mohawk noun incorporation is actually not a feature of the construction as a whole, but of individual noun and verb stems: some stems occur exclusively in such constructions, some often, some occasionally, some rarely, and some never. The situation is similar to that of derivation: productivity is not a feature of derivation as a whole in a language, but rather of individual derivational processes.

Yup’ik shows a similar pattern. The productivity of the verbalizing derivation can be seen in its application to loanwords, a process that occurs more frequently in Yup’ik than in Mohawk. The noun potlatch (a kind of feast), for example, came into the language from Chinook Jargon, which adopted it from Nuuchahnulth patlač ‘gift’.

(46) Yup’ik productivity: loanword base: Elena Charles, speaker p.c.
*Potlatch-aliyalrit.
they potlatch-participate in
‘They come to the potlatch.’

The productivity of the Yup’ik construction is also a property of individual morphemes rather than the process as a whole. Some suffixes are highly productive, such as -u- ‘be’ and -li- ‘make’. Some are marginally productive, such as -ir- ‘deprive’. Some are
unproductive, such as -llite- ‘encounter’. Much is of course domain-specific, such as .-virte- ‘go to’, which occurs only with demonstrative bases that have allative forms, like kia-ni ‘upriver, inside, inland’, kiavirtuq ‘he is going upriver, inside, inland’; (Jacobson 1984: 458, 461).

9. Lexicalization and institutionalization

Word-formation processes function to create lexical items, which can be learned and accessed as units. The property of being stored in memory is often termed lexicalization. Bauer proposes a finer distinction between those stored items which are semantically regular and those which are not.

It might in addition be useful to have a label for those established words which, despite their being established, still form part of a synchronically productive series, differing only from potential words in that, by being used, they have come to have a specific reference. Following Bauer 1983, I shall term such words institutionalized. Institutionalized and lexicalized are thus, by definition, complementary terms, co-hyponyms of established. Bakken (1998: 72) prefers the term conventionalization here, defining a conventionalized word as one which can be recognized and understood out of context, while a coinage requires its context to be comprehensible. She also notes that conventionalization is a scale, sliding into lexicalization.

(Bauer 2001: 46).

The Kapampangan constructions show little evidence of either lexicalization or institutionalization, no more than other syntactic verb-complement constructions. They are generally transparent semantically. The language contains other word-formation devices that serve some of the functions of the incorporating constructions in Mohawk and Yup’ik. There are highly productive derivational verbalizing processes which operate not only on native roots but on borrowed forms as well.

(47) Kapampangan word formation: Clemente Roman, speaker p.c.
Mig-apartmentya.
    mig-apartment=ya
    VERBALIZER-apartment=3SG.ABS
‘She rented her own apartment.’

Mohawk noun incorporation, though much more frequent in speech than its Kapampangan counterpart, shows comparatively strong institutionalization and lexicalization. Speakers are generally aware of which combinations exist in the language and which do not but could, even where all formations are perfectly regular. Awareness of innovations varies to some degree with the individual productivity of the constituents.

There is good evidence that most compound stems are established, stored in memory rather than formed online during speech. The passage in (48) contains
4 clauses, each with an incorporated noun. Each pertains to fish. But the noun root -itsi- ‘fish’ was incorporated in only two of the verbs. The noun root -ia’t- ‘body’ was incorporated in the other two.

\textit{Tanon' ni: kon'takie' ionwanitsiatorà:ton}  
and we all day long we were fish-hunting  
‘And we were fishing all day long  
\textit{kwøh iah ne skaià:ta teionkwentsiaiontà:on.}  
even not that one did we fish-get  
and we didn’t even get one fish.  
\textit{wàksheia'tatshén:ri' karón:token kontià:ti kéntsion'.}  
you body found them log inside they’re body in it fish  
[How could it be that] you found the fish inside a log?’

The speaker was not assembling the verbs with incorporated nouns as he spoke. He was accessing known compound verb stems as units: ‘fish-hunt’, ‘fish-get’, ‘body-find’, and ‘body-be-in’.

Yup’ik shows similar patterns of lexicalization and institutionalization. The same derived forms tend to recur in speech. Speakers know that neqe-te- ‘fish-catch’ exists, but that the combination *luquuyar-te- ‘pike-catch’ does not, even though it would be perfectly regular. Catching pike can be expressed only in separate words: \textit{Luquuyanek pitnaurtutukit.}

As is typical of compounds, the meanings of the Yup’ik derivational constructions are not necessarily semantically compositional. The derived stems are usually formed for a specific purpose and learned and stored as units. They are free to shift in meaning through use, independently of their parts. The suffix -te- does not always contribute the precise meaning ‘catch’: \textit{mallu} ‘beached carcass’, \textit{mallutuq} ‘he found a beached carcass’; \textit{yuk} ‘person’, \textit{yugtuq} ‘he committed murder’; \textit{kass’aq} ‘white person’, \textit{kass’artuq} ‘he went to the city on a shopping trip’ (Jacobson 1984: 586).

In Yup’ik, as in Mohawk, speaker awareness of innovations varies with the productivity of individual morphemes. New formations with highly productive suffixes may be noticed barely if at all, but others are recognized as neologisms even when regular in form and meaning.

10. Conclusion

The Kapampangan, Mohawk, and Yup’ik constructions share some fundamental formal and functional features. All involve the combination of a verbal element and a nominal element to form a new verb headed by the verbal element. All could be classified as endocentric and synthetic. All show a range of semantic relationships between
the incorporated noun and the verb. All provide speakers with choices: verbs with alternative argument structures that can be selected to shape coherent discourse.

The constructions in the three languages show different behaviors with respect to the No Phrase Constraint and the Lexical Integrity Hypothesis, however. The Kapampangan =ng construction often violates the No Phrase Constraint: the nominal elements in these constructions can be syntactically complex. The same construction is also frequently out of line with the Lexical Integrity Hypothesis: the nominal elements in these constructions are apparently visible to the syntax, in that they often serve as antecedents of subsequent pronouns. When the nominal constituent is more tightly bound, however, simply juxtaposed, subsequent reference is not possible. Noun incorporation in Mohawk, by contrast, shows no exceptions to either hypothesis. The incorporated noun is never syntactically complex, and it does not serve on its own as an antecedent for subsequent reference in spontaneous speech. The Yup’ik construction is much like Mohawk noun incorporation. The nominal element may be a noun root or stem, but it is never syntactically complex. It does not normally serve as the antecedent of subsequent pronominals, though in constructions with certain light verbalizing suffixes, it may. The differences among the constructions in the three languages are not random. They correlate with the degree of formal integration of the constituents and the status of the output.

The Kapampangan constructions are the most syntactic: they involve the combination of whole words and phrases. They are not distinguished phonologically from other syntactic constructions. Speakers may pause between the constituents and even insert certain particles. They are fully productive. They do not generally form new lexical items, at least no more often than basic predicate-argument combinations. They are semantically compositional. They can be fully described in terms of general principles: any nominal phrase may be expressed as a non-argument by means of the linker =ng or simple juxtaposition in place of a determiner.

The Mohawk and Yup’ik constructions, by contrast, are solidly word-formation devices. They create single phonological and grammatical words. They generally produce lexical items that are stored and retrieved as units. Their productivity is variable, a characteristic not of the constructions but of the individual morphemes that enter into them. The constructions also vary in their semantic transparency. As institutionalized lexical items, they are free to evolve in meaning with use, independently of the meanings of their constituents. It is important to note that there is no clear line between fully productive incorporating or derivational processes on the one hand, with fully transparent semantics and regular syntactic consequences, and an irregular residue on the other. As word-formation processes, they are exploited to form lexical items corresponding to specific concepts. Productivity of a word-formation process does not necessarily entail full regularity of form and meaning.

Incorporation is thus not a homogenous process, nor is it static over time. The constructions examined here appear to reflect different stages of development. The Kapampangan constructions are typical of early stages: more syntactic and less
integrated. They operate over full words and syntactic phrases, and their constituents are still accessible to speakers. Though they are not core arguments, the nominal elements in =ng constructions can still be referential in their own right and serve as antecedents of subsequent pronouns. The nominal elements of constructions without the linker =ng, more tightly bound to their verbal heads by simple juxtaposition, do not serve as antecedents. It is likely that Mohawk noun incorporation developed from a similar structure, in which uninflected nouns were juxtaposed to verbs as qualifiers rather than syntactic arguments. As the constructions evolved over time, recurring combinations came to be processed as chunks, the formal and semantic bonds grew stronger, and the process became increasingly morphological. The Yup'ik derivational construction is likely to have developed from a stem compounding construction much like that in Mohawk (Mithun 1997, 1998a,b). Certain verb roots that recurred frequently as heads of noun-verb compounds became more general and abstract in meaning over time and began to erode phonologically. The result is a continuum, as formal and functional properties emerge and fade.