KANIEN'KÉHA NOUN INCORPORATION:
A DICHOTOMOUS ANALYSIS*

Martin Renard
University of Toronto

1.   Introduction

Kanien'kéha (Mohawk) is a Northern Iroquoian language spoken in the eight communities of the Kanien'kehá:ka (Mohawk) Nation, across southern Ontario, southern Quebec, and upstate New York. It is severely endangered, with fewer than 700 speakers remaining; although ongoing revitalization efforts are resulting in a growing number of L2 speakers, and the progressive re-establishment of intergenerational transmission (DeCaire 2023).

Like many Indigenous languages of North America, Kanien'kéha forms a perfect example of a "polysynthetic" language, characterized by a rich verbal morphology (Mithun 1999). Polysynthesis as a typological label remains ill-defined, but one of the features with which it is perhaps most uncontroversially associated is the phenomenon of *noun incorporation* (NI), which is pervasive in such languages. Kanien'kéha is no exception.

Descriptively, NI corresponds to the co-occurrence of two lexical roots, typically one nominal and one verbal, in a single phonological word. Consider the following examples:[1]

(1) a.   wa'-ke-nakt-a-hní:non-'
            FAC-1SG.A-bed-LK-buy-PUNC
            'I bought a bed.'

     b.   wake-'nikonhr-a-ién:ta-'s
            1SG.P-mind-LK-obtain-HAB
            'I understand.'

We know that these are single phonological words for multiple reasons which are beyond our scope (Sadock 1980). I assume that all NI forms in this work are single words.

[1] All examples follow the orthography of the Eastern Kanien'kéha dialect, the variety spoken by Kanáhstatsi Nancy Howard, who provided all the data for this work. The spelling system is largely phonemic, and all symbols have their standard IPA values, except for <'> which marks the glottal stop /ʔ/, <i> which also marks the glide /j/ pre-vocalically, and <en> and <on> which mark the nasal vowels /ʌ̃/ and /ũ/ respectively. The following abbreviations are used: 1 = first person; 2 = second person; 3 = third person; A = agent; DUP = duplicative; FAC = factual; FUT = future; FZ = feminine-zoic; HAB = habitual; LK = linker vowel; M = masculine; NMZ = nominalizer; NSF = nominal suffix; P = patient; PUNC = punctual; REL = relative pronoun; REV = reversive; SG = singular; SRFL = semi-reflexive; and STAT = stative. The LK linker vowel is a traditional Iroquoianist label referring to a purely phonologically-governed epenthetic segment, and is not a nominalizer.

The forms in (1) are canonical examples of NI in Kanien'kéha, where a root that looks nominal ('bed', 'mind') and a root that looks verbal ('buy', 'obtain') co-occur in a single word. There is, however, a crucial difference between these two forms: While the meaning of (1a) is compositional, that of (1b) is idiomatic. In fact, all Kanien'kéha NI forms fall neatly into one of these two types, which seem equally important in speech. Throughout this work, I refer to forms of type (1a) as *active*, and those of type (1b) as *inactive*.

This paper addresses the following question: *What are the differences between active and inactive NI forms in Kanien'kéha, and how can we account for them?* I argue that all differences between active and inactive forms derive from the fact that the incorporated element (IE) is syntactically independent in active forms, but not in inactive forms; which can be easily explained by positing additional functional structure in active IEs. Assuming Distributed Morphology (DM; Halle and Marantz 1993), I specifically argue that this extra functional structure is an nP categorization layer. The contributions of this work are the establishment of the hitherto unnoticed empirical dichotomy between active and inactive forms, and the first unified syntactic analysis of *all* NI forms in Kanien'kéha, including idiomatic inactive ones, which have previously often been dismissed as lexical exceptions.

The rest of this paper proceeds as follows. In section 2, I review the debate between the syntactic and lexical approach to NI. Section 3 then presents the categorization-based analysis which forms the core of this work. In sections 4, 5, and 6, I provide evidence for this analysis based on meaning, referentiality, and argument structure, respectively. Section 7 then shows that the predictions of this analysis are verified. In section 8, a final difference between the two types regarding the availability of periphrastic variants is left unexplained, and is accounted for by "excorporation" in section 9. Finally, section 10 concludes.

2.   A longstanding debate

Since Sapir's seminal 1911 article entitled *The problem of noun incorporation in American languages*, NI has received much attention in the literature, and has been considered special due to its ambiguous status as a syntactic-looking word-formation process. This has led to a major debate, which is still alive today: *Is NI syntactic or lexical?*

Major proponents of the lexical approach to NI include Mithun (1984, 1986, 2009), DiSciullo and Williams (1987), and Rosen (1989). In this view, NI stems are derived pre-syntactically inside the lexicon, via morphological compounding of a nominal and a verbal root (VR), often resulting in idiomaticity. Each NI stem forms an independent lexical entry, and is "learned by speakers as a unitary lexical item" (Mithun 1984: 880). NI stems have no syntactically relevant internal structure, and IEs have no syntactic independence.

This lexical view is often pitted against a syntactic approach, for instance developed by Sadock (1980, 1985, 1986), Baker (1988, 1996, 2009), and Barrie and Mathieu (2016). In this approach, NI forms are directly constructed by the generative syntactic component, which merges independent roots during the derivation, resulting in compositional forms. This can occur via head-movement (Baker 1996) or phrasal movement (Barrie and Mathieu 2016). NI stems have a complex internal structure, and IEs are syntactically independent.

At stake in this debate is the larger question of the division of labour between syntax and the lexicon in the derivation of morphologically complex words. This issue can be seen

in the tension between the traditional Lexicalist Hypothesis (Chomsky 1970), whereby words are formed by a separate lexical module and the syntax cannot access their internal structure; and the so-called Single-Engine Hypothesis of more recent frameworks like DM (Halle and Marantz 1993, Barrie 2012), whereby a single computational engine generates all complex structures, including words, which have no special status.

The issue is that this debate is often reductionist. The lexical view denies the syntactic nature of NI, and accordingly focuses on idiomatic inactive forms as representative of the entire phenomenon, thereby extending the assumption of no internal structure to even highly transparent stems. Within the syntactic view, generativists often exclusively examine the most compositional active forms, upholding them as clear evidence of the syntactic nature of NI. Inactive forms are immediately dismissed as "lexicalized" exceptions, with no internal structure worth studying as part of a syntactic theory of NI.

On the one hand, the lexical approach manages to integrate within its purview all forms, including inactive ones, but does so at the expense of the undeniable insight that NI forms are syntactic. On the other hand, the syntactic approach manages to maintain this key idea, but does so at the expense of the study of inactive forms, which are not even considered as instances of "true" NI. This raises a question: How can we both provide a unified account of all Kanien'kéha NI forms *and* maintain their syntactic nature?

I believe that DM (Halle and Marantz 1993) can provide an answer to this question. If we redefine "syntactic" as merely "structural" in a broad DM sense (i.e., composed of multiple sub-units that somehow merge) rather than as "compositional" or "transparent", then it becomes clear that *all* NI forms are inherently syntactic (as they contain multiple distinguishable morphemes) and thus require a syntactic analysis; even inactive ones. This simple but crucial redefinition allows us to ask more seriously of all NI forms, whether they are active or inactive, the truly key question: *What is their internal syntactic structure?*

3.   A categorization analysis

That all Kanien'kéha NI forms are syntactic does not entail that they form a natural class amenable to a single analysis. Indeed, the contrast between *active* and *inactive* NI forms mentioned above is empirically real. The point is that this empirical contrast potentially corresponds to a theoretical contrast between two distinct underlying structures. The goal of this section is to uncover these structures.

We have seen that a major difference between active and inactive forms is that the former are compositional while the latter are idiomatic. In the following sections, we will see that this aligns with three other empirical differences: (i) Active IEs can be referential, but not inactive IEs; (ii) active forms are systematically transitive with the IE functioning as the theme, while inactive argument structures are extremely variable; and (iii) active forms can have periphrastic equivalents, while inactive forms cannot. Arguably, these four differences all derive from a single underlying contrast: The IE is *syntactically independent* in active forms, but not in inactive forms. How can we account for this contrast?
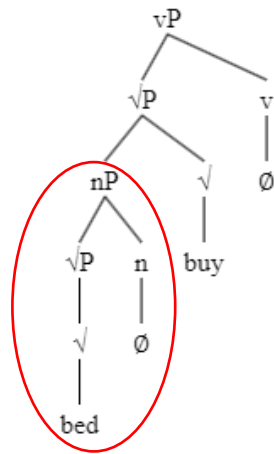
A common intuition is that, if an object is more independent, it is because it contains more functional structure which provides it with this ability. Thus, if active IEs are more independent than inactive IEs, it is because they contain more functional structure than

inactive IEs. What is this additional functional structure that active IEs contain? I make the following proposal: *Inactive IEs are bare uncategorized roots without functional structure, while active IEs contain an nP projection which categorizes their root as nominal*.

One question remains, however: What is the position of active and inactive IEs within their respective syntactic structures? To this point, I make the following concrete proposal: *Inactive IEs are generated as part of a complex head with the VR, while active IEs are generated as the complement of the VR in direct object position*. This follows from the categorization contrast above, as we expect that bare uncategorized roots (inactive IEs) cannot occupy argument (direct object) positions, while categorized roots (active IEs) can.

I thus propose the following structures for active and inactive forms like those in (1) (the position of the IE in each structure is highlighted by a red oval):

(2) a. Active structures   b. Inactive structures
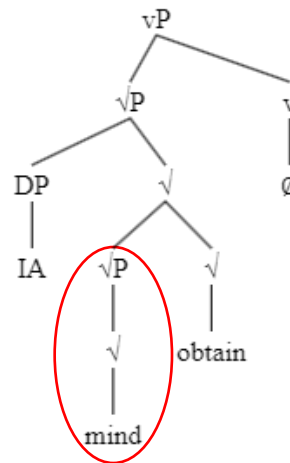
I am only representing vP in this work, as structures above it are irrelevant to us. In active structures like (2a), we have a nominalized IE in complement position; while in inactive structures like (2b), we have a bare uncategorized IE in a complex head with the VR. Note that inactive structures contain an internal argument (IA), and that the n and v categorizing heads are often phonologically empty in Kanien'kéha (although see section 7).

The obvious DM assumption is that roots are not inherently categorized, but merge with categorizing heads (Marantz 1997). Following Harley (2014), I also assume that roots merge with their complement before categorization, and project to √P. Finally, I assume that there is a structural asymmetry in inactive forms, whereby it is the VR rather than the IE that projects to the head √P. This shows that, although inactive stems have no *functional* structure, they do have *internal* structure. Evidence for this includes the fact that it is the VR which controls the selection of the agreement series, but this is beyond our scope.

These structures are not brand new. Barrie and Mathieu (2016) similarly argue that NI in Northern Iroquoian languages targets nP, but do not recognize a contrast between nP and √P NI, presumably because they also dismiss inactive forms as non-syntactic. As for the position of the IEs, Baker (1996) proposes that they are generated in argument position, as in (2a), and then head-move to the verbal head, as in (2b). Of course, in this analysis, I assume no derivational relation between (2a) and (2b); these are just different NI types.

Finally, this analysis of inactive forms resembles Harley's (2009) treatment of compounds in DM, with two bare roots in a direct sisterhood relation yielding an idiomatic meaning.

4.   Meaning

Let us now motivate this categorization analysis further by examining how it accounts for the various empirical contrasts between active and inactive forms, starting with their meaning. On the one hand, we have seen that active forms are always compositional:

(3)   wa'-ke-nénhst-a-ke-'
        FAC-1SG.A-corn-LK-eat-PUNC
        'I ate corn.'

Here, the meaning of *-nenhst-* 'corn' and that of *-ke-* 'eat' combine compositionally to form the meaning of the whole stem *-nenhstake-* 'eat corn' in a predictable way.
        On the other hand, all inactive forms tend to have an idiomatic interpretation:

(4)   te-ke-rihw-á-hkhw-a'
        DUP-1SG.A-matter-LK-pick.up-HAB
        'I sing.'

Here, the meaning of *-rihw-* 'matter' and that of *-hkhw-* 'pick up' combine unpredictably to form the idiomatic meaning 'sing' for the whole stem *-rihwahkhw-*.
        How, then, does our categorization analysis account for these patterns? A common assumption is that categorizing heads, as the first elements which roots merge with, systematically function as phase heads and trigger spell-out to the interfaces LF and PF of the whole structure which they delimit (i.e., the categorization projection that they head itself, and everything merged under it) (Marantz 2000, Arad 2003, Embick and Marantz 2008). As shown in (2a), active IEs correspond to whole nPs dominating a √P. Because nPs are headed by n, which is a categorizing head and thus a phase head, the whole nP structure that they delimit (i.e., the whole active IE) is spelled out and independently interpreted at LF first, before any further computational operation. This means that by the time the active IE merges with the VR, its meaning has been spelled out and cannot be further semantically manipulated, necessarily resulting in a compositional interpretation.
        Inactive IEs, on the other hand, contain no categorizing head. This means that no phasal boundary intervenes between the IE and the VR with which it merges, and that both are simultaneously spelled-out and interpreted at LF together as a single phase when the derivation reaches the first phase head. In this case, this is the v categorizing head, which triggers spell-out of the whole vP structure it delimits. Undissociated spell-out in the same phase allows for a tighter and more idiosyncratic negotiation of meaning between the two roots, which may often result in an idiomatic meaning. Obviously, this does not mean that the resulting meaning is random. Rather, as a bundle of two lexical roots is spelled-out without any functional structure to drive a compositional interpretation, the Encyclopedia

(assuming a classical DM framework) supplies a special idiomatic meaning based on the identity of the two roots, which in a sense are interpreted in the context of each other.

Some forms are ambiguous between a compositional and an idiomatic meaning:

(5) ro-nia'tar-á:-k-on
3SG.M.P-cloth-LK-eat-STAT
(i) 'He is eating cloth.' (compositional)
(ii) 'He complains a lot.' (idiomatic)

This NI stem *-nia'tarak-* is ambiguous between a compositional meaning 'eat cloth' and an idiomatic meaning 'complain a lot'. I assume that the two meanings in such ambiguous NI stems derive from a structural ambiguity between two competing underlying structures, namely one active with a categorized IE yielding a compositional interpretation, and one inactive with a bare uncategorized IE yielding an idiomatic interpretation.

5.   Referentiality

An effective method to test for the referentiality of an IE is to determine whether the IE can be modified by external modifiers, that is modifiers located outside of the NI form (Baker 1996). If an IE can be modified by an external modifier, it must be referential, as it is the morpheme that introduces the modified referent. Inversely, if an IE cannot be modified by an external modifier, then it is not referential, as it fails to introduce a referent for modification. Relative clauses are a good type of modifiers to use, as the anaphoric relation between the antecedent and the relative pronoun is local, thereby reducing the risks of interference from pragmatically determined anaphoric relationships. In Kanien'kéha, relative clauses are introduced by the collocation *tsi niká:ien* 'which, the one which'.

When this relativization test is applied to Kanien'kéha IEs, a clear pattern arises. On the one hand, active IEs can be modified by external relative clauses:

(6) wa'-ke-nonhs-a-hní:non-'   tsi niká:ien   se-nòn:we-'s
    FAC-1SG.A-house-LK-buy-PUNC   REL   2SG.A-like-HAB
    'I bought the house that you like.'

The external relative clause *tsi niká:ien senòn:we's* 'that you like' modifies the IE *-nonhs-* 'house', which thus functions as the antecedent of the relative pronoun *tsi niká:ien* 'that'. This suggests that this IE is referential, in the sense of introducing the new referent *house* into the context, which otherwise would not be available for external relativization. All active forms yield similar results, and thus seem to contain a fully referential IE.

On the other hand, inactive IEs cannot be modified by external relative clauses:

(7) wake-'nikonhr-a-ién:ta-'s   tsi niká:ien   se-nòn:we-'s
    1SG.P-mind-LK-obtain-HAB   REL   2SG.A-like-HAB
    'I obtain the mind that you like.' (Intended: unclear)

Trying to externally modify an inactive IE always forces a compositional reading, as the IE has to be made independent (i.e., active) to be externally relativized, so that it is not the inactive IE itself that is being modified. The resulting reading may often be nonsensical, as is the case here, because inactive forms do not always have a semantically acceptable literal interpretation. In these cases, a more acceptable meaning in which the relative clause modifies an implied internal argument ('I understand what you like') may be available, but then it is again not the inactive IE itself that undergoes modification. In any case, the point is that no inactive IE can be externally modified *as such* (and it is unclear what meaning this would even create), suggesting that no inactive IE is able to introduce referents.

In ambiguous forms, relativization is possible, but always results in an active reading:

(8) ro-nia'tar-á:-k-on   tsi niká:ien   se-nòn:we-'s
     3SG.M.P-cloth-LK-eat-STAT   REL   2SG.A-like-HAB
     'He is eating the cloth that you like.'

In this ambiguous form repeated from (5), only the compositional meaning is available with the external relative clause, again because the IE has to be made independent (i.e., active) to be externally relativized. The idiomatic meaning is unavailable with this external relative clause; in fact, it is not even clear what it would be. This confirms that active IEs can be externally modified (and hence are referential), while inactive IEs cannot be externally modified (and hence are not referential). This makes sense, as a nominal can only be externally modified if it is syntactically independent (i.e., active), and not if it is idiomatically interpreted as part of a non-compositional expression (i.e., inactive).

Thus, contrary to previous lexical approaches claiming that IEs are *never* referential (e.g., Mithun 1984), or previous syntactic approaches claiming that IEs are *always* fully referential (e.g., Baker 1996), we have seen that it simply depends on the type of the form: In active forms, the IE is referential; while in inactive forms, it is not. How, then, does our categorization approach account for this contrast in referentiality?

A common intuition is that, if a syntactic object can refer, then it must contain some functional structure responsible for referential import; and Barrie and Mathieu (2016) suggest that the nP projection can fulfill this function. Active IEs, which contain this projection, are thus able to refer; while inactive IEs do not contain any functional structure, and are thus defective with respect to referential ability. A broader cross-linguistic question is which nominal projection is responsible for which kind of referentiality, but it is sufficient for our purposes to have a contrast between *some* functional structure, granting referentiality to active IEs, and *no* functional structure, making inactive IEs unable to refer. In this analysis, I argue that this functional structure is the categorization projection nP.

6.   Argument structure

We have so far been providing evidence for the proposed *internal structure* of IEs: Active IEs are referential and are interpreted compositionally, confirming that they contain the functional layer nP; while inactive IEs are not referential and are interpreted idiomatically

with the VR, confirming that they contain no functional structure. What evidence can we provide to support the proposed *structural positions* of IEs, however?

Such evidence can be found in argument structure. Consider the examples below:

Table 1. Argument structures in active Kanien'kéha NI forms

IE	VR	NI stem	Transitivity	Θ-role of IE
-nakt- 'bed'	*-hninon-* 'buy'	'buy a bed'	Transitive	Theme
-nonhs- 'house'	*-onni-* 'build'	'build a house'	Transitive	Theme
-nenhst- 'corn'	*-k-* 'eat'	'eat corn'	Transitive	Theme
-ient- 'wood'	*-ontho-* 'burn'	'burn wood'	Transitive	Theme

Table 2. Argument structures in inactive Kanien'kéha NI forms

IE	VR	NI stem	Transitivity	Θ-role of IE
-'nikonhr- 'mind'	*-ienta-* 'obtain'	'understand'	Transitive	Goal (?)
-'nikonhr- 'mind'	*-hen-* 'fall'	'forget'	Transitive	Source (?)
-hah- 'road'	*-hkhw-* 'pick up'	'walk'	Unergative	Location (?)
-hah- 'road'	*-ionni-* 'extend'	'be Wolf Clan'	Unaccusative	Theme (?)

Table 1 shows that all active forms systematically have the same argument structure, irrespective of the identity of the IE or the VR: The NI form is a simple transitive stem, in which the IE is interpreted as the theme. On the contrary, in Table 2, we can see that, even if the IE remains constant, the argument structure of each inactive form varies between transitive, unergative, and unaccusative, and the thematic relationship between the IE and the VR is unclear, so that attempts at using common labels such as "goal" remain uncertain.

How do the different positions of active and inactive IEs explain this contrast in argument structure? In active forms, the IE is always the complement of the VR. As this is the position of the internal argument or direct object, the IE is always interpreted as the theme of the VR, following Baker's (1988) Uniformity of Theta Assignment Hypothesis. Given that Kanien'kéha verbs always require an agreement marker (irrespective of the presence or absence of an IE), which here must realize an external argument as the internal argument position is occupied, we systematically obtain a transitive construction (i.e., a construction with one external and one internal argument). Unaccusative active forms, in which the agreement marker co-indexes the internal argument realized by the IE, seem possible on the surface in a few cases, but in fact a closer examination reveals that these are not true instances of active NI; although this is beyond our scope.

In inactive forms, the IE is part of a complex verbal head, and is not an argument at all, so that all argument positions are free. This predicts that there should be no restrictions on the range of possible argument structures, which is confirmed by the variability in Table 2. Moreover, these verbal heads are composed of two bare uncategorized roots that merge without any functional structure intervening between them, yielding a much freer and more undefined thematic relation. This underdetermines thematic interpretations, and eludes the

labels used for the more systematic thematic roles of syntactic arguments. The lack of a mediating functional layer that could have driven a more predictable semantic composition and thematic relation means that there is "no consistent syntactic or even semantic relation between incorporated nouns and the verb roots that incorporate them" (Mithun 2009: 575), so that "it is up to the interpretive component to construct some plausible relationship between the incorporated noun and the [VR; MR]" (Harley 2009: 139).

7.   Predictions

We have seen that the categorization analysis proposed in section 3 is supported by major empirical contrasts between active and inactive forms, which this analysis readily explains. To fully motivate this approach, however, we need to verify its predictions.

   The major structural difference between active and inactive forms is the presence of an nP projection in active IEs. Kanien'kéha independently features overt nominalization markers, which can be analyzed as morphological realizations of the n nominalizing head. Assuming that n can also be phonologically empty in some contexts, as is common cross-linguistically, our analysis makes two concrete predictions: (i) In active forms, there will *sometimes* be an overt nominalizer between the IE and the VR (because there is an underlying n head); (ii) in inactive forms, there will *never* be an overt nominalizer between the IE and the VR (because there is no underlying n head).

   Let us begin with the first prediction. In Kanien'kéha NI forms, an overt nominalizer must always be inserted when the IE is either (i) morphologically verbal or (ii) a loanword (the overt nominalizers and their glosses are bolded for clarity):

(9)   a.   wa'-ke-'sere-**ht**-a-hní:non-'
           FAC-1SG.A-drag-**NMZ**-LK-buy-PUNC
           'I bought a car.'

      b.   en-wak-job-**hsher**-a-ién:ta-'ne'
           FUT-1SG.P-job-**NMZ**-LK-obtain-PUNC
           'I will get a job.'

In (9a), the VR -*'sere*-, which is idiomatically used for 'car', is incorporated, triggering nominalization. In (9b), the loanword -*job*- 'job' is incorporated, which also triggers nominalization. Without an overt nominalizer, these forms would be ungrammatical.

   Note that two different allomorphs of the nominalizer are used, due to different diachronic origins; but these synchronically fulfill the same function. In these cases, nominalization is presumably triggered by a syntactic constraint whereby only roots categorized as nominal may incorporate, so that roots that are usually verbal, as in (9a), or non-native roots without a category, as in (9b), need to be licensed through nominalization.

   Crucially, both forms in (9) are active, as shown for instance by the compositional combination of the meaning of the IE and that of the VR (e.g., 'car' + 'buy' = 'buy car'). Even though it is not directly the property of being active, but rather that of containing a verb or a loanword, that triggers overt nominalization, this means that the first prediction

is verified: In active forms, there is *sometimes* an overt nominalizer between the IE and the VR. This provides direct empirical support for the presence of an underlying n head in active forms, and therefore also for the presence of an nP projection in active IEs.

Let us move on to the second prediction. An examination of a large inventory of inactive forms reveals that none contain an overt nominalizer. Moreover, there are cases where nominalization is expected for independent reasons, but fails to occur:

(10) wak-at-onhnh-á-her-e'
      1SG.P-SRFL-live-LK-be.on.top-STAT
      'I am thrilled.'

Kanien'kéha sometimes allows the incorporation of VRs like *-onhnh-* 'live' without overt nominalization, as in (10). In fact, in this case a nominalizer would be ungrammatical.

The form in (10) is clearly inactive, as it is idiomatic (i.e., 'live on top' = 'be thrilled'). The point is that, even though the IE is verbal (and as such normally requires a nominalizer, as we have seen for active forms above), overt nominalization is impossible, arguably *because* the form is inactive. This suggests that the second prediction is also verified: Inactive forms never contain an overt nominalizer, even in the presence of independent constraints which normally trigger one in active forms. This supports our analysis of inactive IEs as lacking a functional categorization projection nP headed by n.

Given the structures in (2), another prediction is that, beside the n head in active forms, IEs are monomorphemic; that is, a single atomic root is incorporated. As notoriously argued by Barrie and Mathieu (2016), this is a dangerous empirical over-simplification:

(11) wa'-k-**at-o-kwa**-hsher-óhare-'
      FAC-1SG.A-**SRFL-immerse-REV**-NMZ-wash-PUNC
      'I washed a spoon.'

In this example, a complex verbal structure (bolded) rather than a root is incorporated. This could be taken as evidence against our analysis, but I argue that it is not. The form in (11) is active, as suggested by its compositional meaning (i.e., 'spoon' + 'wash' = 'wash a spoon'). In fact, all forms containing a morphologically complex IE seem active. In such cases, we could simply argue that a complex structure such as vP rather than a bare √P is nominalized by merging with n. In inactive forms, however, it would be difficult to argue that a vP merges inside a complex head, so it seems significant that no such forms exist.

8.   Periphrastic variants

We have so far focused exclusively on NI, in order to account for the distinction between active and inactive NI. However, NI forms are often contrasted with periphrastic variants:

(12) a.   ka-nákt-a'   wa'-k-hní:non-'
            3SG.FZ.A-bed-NSF   FAC-1SG.A-buy-PUNC
            'I bought a bed.'

b. wa'-k-hní:non-'   ne   ka-nákt-a'
    FAC-1SG.A-buy-PUNC   NE   3SG.FZ.A-bed-NSF
    'I bought a bed.'

c. ?o-'nikòn:r-a'   wak-ién:ta-'s
    3SG.FZ.P-mind-NSF   1SG.P-obtain-HAB
    'I obtain a mind.' (Intended: 'I understand.')

In (12a) and (12b), we can see that active NI forms like (1a) can take truth-conditionally equivalent non-NI forms, and that these can be of two major types, depending on whether the external nominal (EN) is pre- or post-verbal. If post-verbal, the EN is systematically preceded by the particle *ne*, which we will discuss below. In (12c), however, we can see that inactive NI forms like (1b) cannot take truth-conditionally equivalent non-NI forms. Trying to construct one forces a compositional meaning (which may be non-sensical, as we have seen), so that the inactive idiomatic meaning is lost. Similarly, non-NI equivalents of ambiguous forms like (5) are always interpreted compositionally.

This raises several questions: (i) What is the structure of non-NI forms?; (ii) What is the relationship between NI and non-NI forms?; and (iii) Why do active forms have non-NI equivalents, but not inactive forms? Clues to answer these questions can be found in the discourse distribution of NI vs. non-NI in Kanien'kéha. As DeCaire et al. (2017) show, NI (where it is morphophonologically available) is in fact the obligatory unmarked default in most contexts, while non-NI is only licensed in marked information-structural contexts.

Thus, forms like (12a) only occur in cases where focus (Rooth 1992) falls on the object, which is thus fronted instead of being incorporated. Similarly, forms like (12b) only occur in cases where focus falls on the verb, which forces the object to occur in final position, so that the verb may be isolated and fronted. Non-NI forms occur nowhere else. I assume that, in cases of verb focus, the object is relegated to antitopic status in the right-periphery, which is marked by *ne*. This is because post-verbal nominals introduced by *ne* conform to the definition of antitopics: They are entirely in the *ground*, and fully devoid of *figure* (Gordon 2008). Thus, while corresponding active NI and non-NI forms are truth-conditionally equivalent, they are not information-structurally equivalent.

9.   An excorporation analysis

How can we explain these patterns? Contrary to the mainstream analysis whereby object nominals are base-generated outside of the verb and then incorporate into it (Baker 1996), I follow the insights of DeCaire et al. (2017) in arguing for the opposite analysis: All object nominals are base-generated inside verbal complexes as IEs, and only move out of it in marked information-structures; that is, we have *excorporation* instead of incorporation.

While I assume that all NI forms are base-generated, I argue that non-NI forms necessarily derive from underlying *active* NI forms (so that they are always interpreted compositionally), because ENs derive from active IEs. Indeed, ENs always contain a "noun suffix" (NSF), which is arguably the realization of the n head in active IEs. A crucial difference between active IEs and ENs, however, is that ENs contain an additional

agreement marker, which I argue heads a DP projection. This makes sense: ENs are even more syntactically independent than active IEs, so they must contain more functional structure which grants them this independence. We thus have a continuum in Kanien'kéha nominals from ENs (more independent; DP + np + √P), to active IEs (less independent; np + √P), and finally to inactive IEs (not independent; only √P); and this may reflect a diachronic pathway of lexicalization (understood here as the loss of functional structure).

Following Boles (2023), I argue that Kanien'kéha DPs are always generated with information-structural features, and thus trigger excorporation. In cases of object focus, the object nominal is generated with a DP layer which contains a [focus] feature, triggering focus movement from the underlying incorporated position to the focus position Spec-FocP in the left-periphery. In cases of verb focus, the object nominal is generated with a DP layer which this time contains an [antitopic] feature, triggering antitopic movement from the underlying incorporated position to the antitopic position Spec-AntitopP in the right-periphery. Contrary to FocP, the head of AntitopP is overtly realized as the particle *ne*. Antitopic movement isolates the verbal complex, which is then free to raise to Spec-FocP to receive focus. That antitopic movement occurs first before focus movement suggest that AntitopP is generated first below FocP (assuming minimalist bottom-up generation).

The structures of sentences like (12a) and (12b) are thus respectively represented in (13a) and (13b) below (the constituents that move are highlighted with a red oval):
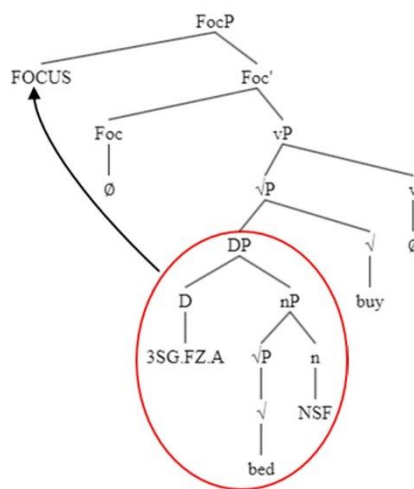
(13) a.   Object focus   b.   Verb focus
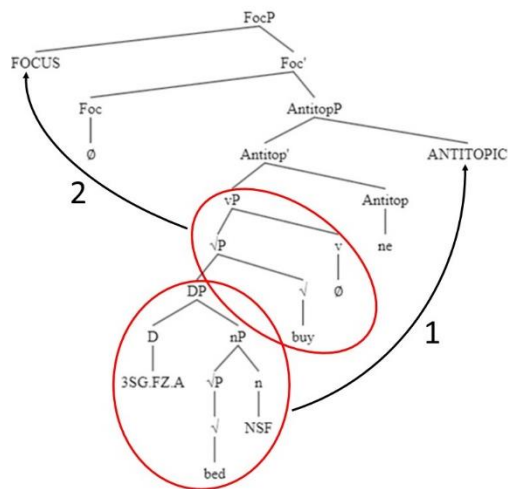
As these structures show, NI (where available) is the underlying structure of all transitive clauses in Kanien'kéha, and non-NI is only derived when the underlying active IE is generated with a DP bearing information-structural features, and thus excorporates to the left-periphery for focus, or to the right-periphery for antitopic, thereby becoming an EN.

This analysis presents several advantages. First, it captures the discourse distribution of NI vs. non-NI, which is missed in Baker's (1996) approach. Second, this view naturally explains why corresponding NI and non-NI forms are only truth-conditionally equivalent, but not information-structurally equivalent; which is missed in Baker's approach where these are considered to be essentially synonymous (1996).

Third, this analysis reflects a key empirical fact: In polysynthetic languages like Kanien'kéha, NI is not a marked structure at all, contrary to what might be suggested by its derived status within theories biased towards analytic Indo-European languages (compared to which NI is bound to appear special). On the contrary, NI is the unmarked and pervasive default in discourse, and it is rather non-NI that is only licensed in marked and restricted contexts. The mainstream Bakerian approach misses this key insight.

Fourth, and finally, this analysis provides a direct answer to our original question. A common intuition is that a constituent requires functional structure to be able to move. Thus, only active forms have non-NI equivalents, because only active IEs have functional structure (nP) and so can merge under a DP triggering their excorporation (assuming DP cannot take bare roots as complements). In inactive forms, however, the IE is a bare root without functional structure, and thus can neither move on its own nor merge under a DP to excorporate, making a non-NI equivalent unavailable. Trying to create the non-NI equivalent of an inactive NI form by artificially adding an nP and DP layer and excorporating the IE forces a compositional meaning, so that the idiomatic meaning is lost.

10. Conclusion

I have made three main claims in this work. First, I have proposed a novel empirically real distinction between two structurally different types of NI in Kanien'kéha, namely *active* NI and *inactive* NI, which had to my knowledge never been fully established before.

Second, I have argued that all differences between active and inactive NI forms follow from the fact that the IE contains a functional categorizing layer nP in active forms, whereas it is a bare $\sqrt{}$P in inactive forms; thereby providing the first unified syntactic analysis of *all* Kanien'kéha NI forms. This therefore shows that, even though they lack *functional* structure relative to active forms, inactive forms do contain *internal* structure; even if only composed of lexical roots. Therefore, all NI forms are *syntactic* in a broad DM sense, even the most idiomatic inactive ones, and there is no need to dismiss them as lexicalized exceptions, like analysts across the theoretical spectrum have done for decades.

Third, I have shown that the information-structurally governed distribution of non-NI forms relative to NI forms in Kanien'kéha discourse suggests a reversal of traditional analyses: NI forms are base-generated, and non-NI forms are derived from them by *excorporating* the noun to focus or antitopic position. This approach captures a key insight that is often missed in theories biased towards non-incorporating Indo-European languages: In polysynthetic languages, NI is not a marked or special construction in any sense, but rather an unmarked and default structure which is pervasive in speech, and which bears a functional load similar to that of basic multi-word clauses in more analytic languages. This complete empirical reversal is thus reflected in the analysis.

We have also seen that these two distinct proposals were nevertheless connected, specifically when it came to the availability of truth-conditionally equivalent non-NI forms: Only active NI forms can have semantically equivalent non-NI forms, because only active IEs contain the functional structure necessary to excorporate and become ENs. These two analyses are also related in a more general way, however: The little functional structure that base-generated IEs contain, whether none at all in inactive forms or merely a single nP

layer in active forms (as per the categorization analysis), may be the reason why NI is so pervasive and hence ENs are so rare in actual discourse in polysynthetic and incorporating languages like Kanien'kéha (as per the excorporation analysis). Indeed, Kanien'kéha nominals are usually generated with only little functional structure (i.e., nP or nothing), which limits their syntactic independence; and they are thus often only licensed as IEs (i.e., base-generated inside the verb). Licensing them as ENs is more costly because it requires the generation of additional functional structure which inherently contains information-structural features (e.g., DP); and this can only occur in marked and rarer information-structural contexts, such as focus (pre-verbal ENs) and antitopic (post-verbal ENs).

Presumably, this constitutes only one end of a larger typological continuum. On this end are verb-heavy languages like Kanien'kéha, in which nominals are typically generated with little functional structure and hence little syntactic independence, and thus start out low in the structure deep inside the verbal domain. It is costly to move them up the structure and outside of the verbal domain, as the additional functional structure required to do so is only licensed under special conditions. On the other end are noun-heavy languages such as Arabic, in which nominals tend to be generated with significantly more functional structure, providing them with greater syntactic independence and allowing them to behave independently from the verbal projection at no extra computational cost. Of course, this remains a speculative hypothesis in need of further research.

This work has focused exclusively on a single language, namely Kanien'kéha, a Northern Iroquoian language spoken in Ontario, Québec, and upstate New York. Although the results and analysis are probably straightforwardly applicable to closely related Northern Iroquoian languages, a natural question to ask here is whether and to what extent these categorization and excorporation approaches can be extended to other unrelated polysynthetic and incorporating languages as well. I would like to predict that it could indeed be the case, but one must never forget that "noun incorporation constructions in different languages seem to be different enough syntactically and semantically to warrant distinct analyses", so that "[i]t may well be that most of the NI theorists are correct for the language(s) they know best, and become wrong only if they say or imply that there is a single unified syntax for all the constructions called noun incorporation in the languages of the world" (Baker 2009: 164). Inversely, although I have been assuming uniformity, more work needs to be done regarding the potential effects of smaller-scale variation (e.g., community, family, or individual dialects) on the discourse usage and syntactic structure of inactive, active, and non-NI forms *within* the Kanien'kéha speech community.

References

Arad, Maya. 2003. Locality constraints on the interpretation of roots. *Natural Language and Linguistic Theory* 21: 737–778.

Baker, Mark C. 1988. *Incorporation: A theory of grammatical function changing.* Chicago: University of Chicago Press.

Baker, Mark C. 1996. *The Polysynthesis Parameter.* Oxford: Oxford University Press.

Baker, Mark C. 2009. Is head-movement still needed for noun incorporation? *Lingua* 119: 148–165.

Barrie, Michael, and Eric Mathieu. 2016. Noun incorporation and phrasal movement. *Natural Language and Linguistic Theory* 34: 1–51.

Barrie, Michael. 2012. Noun incorporation and the lexicalist hypothesis. *Studies in Generative Grammar* 22(2): 235–261.

Boles, Chase. 2023. Accounting for possession in Kanien'kéha (and Oneida). Paper presented at MULL Lab, McGill University, Montréal, QC.

Chomsky, Noam. 1970. Remarks on nominalization. In *Readings in English Transformational Grammar*, ed. Roderick Jacobs and Peter Rosenbaum, 184–221. Waltham: Ginn.

DeCaire, Ryan, Alana Johns, and Ivona Kučerová. 2017. On optionality in Kanien'kéha noun incorporation. *Toronto Working Papers in Linguistics* 39: 1–10.

DeCaire, Ryan. 2023. The role of adult immersion in Kanien'kéha revitalization. Doctoral Dissertation, University of Hawai'i at Hilo.

DiSciullo, Anna-Maria, and Edwin Williams. 1987. *On the definition of word*. Cambridge, MA: MIT Press.

Embick, David, and Alec Marantz. 2008. Architecture and blocking. *Linguistic Inquiry* 39(1): 1–53.

Gordon, Bryan. 2008. Antitopic and afterthought: Givenness and grammar in Ponca and Omaha. *Kansas Working Papers in Linguistics* 30: 74–87.

Halle, Morris, and Alec Marantz. 1993. Distributed Morphology and the pieces of inflection. In *The view from building 20: Essays in linguistics in honor of Sylvain Blomberger*, ed. Kenneth Hale and Samuel Jay Keyser, 111–176. Cambridge, MA: MIT Press.

Harley, Heidi. 2009. Compounding in Distributed Morphology. In *The Oxford Handbook of Compounding*, ed. Rochelle Lieber and Pavol Štekauer, 129–144. Oxford: Oxford University Press.

Harley, Heidi. 2014. On the identity of roots. *Theoretical Linguistics* 40(4): 225–276.

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers in Linguistics* 4(2): 201–225.

Marantz, Alec. 2000. Roots: The universality of roots and pattern morphology. Paper presented at the Conference on Afro-Asiatic Languages, University of Paris VII, Paris, France.

Mithun, Marianne. 1984. The evolution of noun incorporation. *Language* 60(4): 847–894.

Mithun, Marianne. 1986. On the nature of noun incorporation. *Language* 62(1): 32–37.

Mithun, Marianne. 1999. *The languages of Native North America*. Cambridge, UK: Cambridge University Press.

Mithun, Marianne. 2009. Iroquoian: Kanien'kéha. In *The Oxford Handbook of Compounding*, ed. Rochelle Lieber and Pavol Štekauer, 564–583. Oxford: Oxford University Press.

Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1(1): 75–116.

Rosen, Sara T. 1989. Two types of noun incorporation: A lexical analysis. *Language* 65(2): 294–317.

Sadock, Jerrold. 1980. Noun incorporation in Greenlandic: A case of syntactic word formation. *Language* 56: 300–319.

Sadock, Jerrold. 1985. Auto-Lexical Syntax: A proposal for the treatment of noun incorporation and similar phenomena. *Natural Language and Linguistic Theory* 3: 379–439.

Sadock, Jerrold. 1986. Notes on noun incorporation. *Language* 62(1): 19–31.

Sapir, Edward. 1911. The problem of noun incorporation in American languages. *American Anthropologist* 13: 250–282.