Historical Linguistics 1999

EDITED BY
Laurel J. Brinton

Offprint
Lexical forces shaping the evolution of grammar

Marianne Mithun
University of California, Santa Barbara

A feature commonly noted in typological descriptions of languages from early times to the present has been their degree of synthesis, their average number of morphemes per word. At one extreme of the synthesis continuum are analytic constructions, in which most words consist of a single morpheme. An analytic structure is in (1) from Engenni, a Kwa language of Nigeria.

(1) Analytic construction from Engenni (Thomas 1978:121)
á  ta na wa  ñmu
one go to seek house
"Let's go look for the house."

At the other extreme are polysynthetic constructions, in which words typically consist of large numbers of morphemes. An example of a more polysynthetic structure can be seen in (2) from Mohawk, an Iroquoian language of Quebec, Ontario, and New York State.

(2) Polysynthetic construction from Mohawk
	teninonhsisâkha
teni-nonhs-ihsak-ha
1.INCL.AGENT.DUAL-house-seek-ANDATIVE
"Let's go look for the house."

One might wonder whether a construction like that in (2) is really a single word. There are numerous formal indications that it is, including the position of primary stress, the operation of various phonological processes, and its morphological structure, but of special significance are the judgments of Mohawk speakers. Whether or not they have ever given much conscious thought to their language or seen it written, they immediately recognize (2) as a single word. They cannot usually identify its components in isolation. Asked the meaning of -ni- or -nonhs- for example, even the best speakers are at a loss (unless trained as linguists), nor are
they be able to isolate which elements mean "look for" or "go and". The contrast in structure between (1) and (2) raises an obvious question: does it matter how much information is packaged in a word?

There is evidence that it does. Languages with polysynthetic constructions generally offer their speakers analytic constructions as well. Each of the elements of the Mohawk word in (2) could have been expressed with a separate word. The language contains an independent first person pronoun i-}: an independent numeral tékeni "two", an independent nominal kanónhsote "house", and a full verb root -e- "go". Such options would not coexist so systematically alongside of the bound morphemes of (2) if they were functionally equivalent to them. The alternatives do indeed serve different functions. Synthetic constructions are used by Mohawk speakers to package together elements of what is treated as a single idea or concept. Separate words, by contrast, are used to focus special attention on their individual content, such as to introduce significant new information, add emphasis, or highlight a contrast. Similarly when English speakers use the synthetic form eggbeater, they are likely to visualize a familiar tool, rather than first imagining eggs that have been broken, then the action of beating them, then various objects that could be used for the action. The -er suffix is also used in English to derive agentive nominals, but in this case it is the instrument that springs to mind rather than a person hitting eggs with his fists or another tool, because we immediately recognize the word as a lexicalized unit. We might, on the other hand, first introduce an unfamiliar, complex idea with a phrase, such as a “device for registering the serial numbers of bills”. Here we will see that the phenomenon of lexicalization can affect not just language use but language change as well.

In the Northern Iroquoian languages of northeastern North America, just three lexical categories have traditionally been recognized: nouns, verbs, and particles. Qualities expressed by adjectives in other languages are predicated by verbs. The early stages of development of a distinct attributive adjective category can now be discerned, however. The nature of this evolution demonstrates the potentially powerful role of lexicalization in certain processes of grammatical change.

1. Lexical categories

Words in Northern Iroquoian languages fall into three classes on the basis of their internal morphological structure: particles, nouns, and verbs. Particles are by definition morphologically unanalyzable. Examples here are Mohawk, but the structures are the same in the related languages.

(3) Morphological particles

\[ \text{wáhi'} \quad \text{ki'ken} \]

"Isn't that so?"       "this"
Nouns consist of a prefix indicating the gender of the referent or its possessor, a noun stem, and a noun suffix, which contributes no meaning but identifies the word as a formal noun.

(4) Morphological nouns

- orawista'  
- or-ra'wista'
- 1.sg.poss-pancake-noun.suffix

“pancake”

Nominals may also be followed by various word-level suffixes, including locatives, residential, a characterizer, a distributive, an augmentative, a diminutive, and a decessive.

(5) kahnawa'kehrnon'kenha'  
ka-hnaw-a'-ke-hronon'-kenha'

“the former (or late) residents of the place at the rapids”

Verbs consist minimally of a pronominal prefix specifying the core arguments of the clause, a verb stem, and an aspect suffix.

(6) Morphological verb

Riketskwas  
ri-ketska-s

“I wake him up.”

They may also contain a number of other affixes. In addition to the obligatory pronominal prefix, there may be coincident, contrastive, partitive, negative, translocative, factual, duplicative, future, optative, cislocative, repetitive, reflexive/reciprocal, and/or middle prefixes. In addition to the aspect suffixes (perfective, imperfective, or stative), there may be inchoative, reversive, causative, instrumental applicative, benefactive applicative, distributive, andative, facilitative, purposive, continuative, and/or various past tense suffixes.

(7) More complex verbs

a. Tonahshako'nikohrotakwenhákie'  
  t-on-ta-shako-'nikonhr-ot-akw-en-hatie'

“he changes her mind at every turn” = “He rules over her (mentally).”

b. Enkatatekhónminen  
en-k-atate-khw-omni-en-

“I’ll cook for myself.”
Verbs may also contain incorporated noun stems, like ’nikonhr- “mind” in (7a) and -khw- “food” in (7b). Only the noun stem is incorporated: the full noun for “food” is kâ-khw-a, with neuter prefix and noun suffix, but only the stem -khw- appears in the compound verb stem -khw-onni “food-make” = “cook”. The incorporated noun does not function as a syntactic argument. The core arguments of (7a), for example, are "he" and "her", not "he" and “mind”, as can be seen in the pronominal prefixes. Verbs formed with noun incorporation in Mohawk may be transitive, intransitive, or ambivalent. They are formed for a purpose, as labels for nameworthy concepts. Once lexicalized, they are stored, accessed, and learned by other speakers as single units.

Qualities expressed by adjectives in other languages are usually predicated in Northern Iroquoian languages by stative verbs. Like all verbs, they show morphological marking for aspect, tense, negation, and other typical verbal distinctions. Many appear with incorporated nouns.

(8) Mohawk statives with typical verbal morphology

a. kakowd:nen
   ka-kowan-en
   NEUT.1BE.BIG-STATIVE
   “It is big.”

b. Kahentowanen:ne
   ka-hent-owan-en-hne
   NEUT.1FIELD-BE.BIG-STATIVE-PAST
   “It was a big field.”

c. Iah tekanatowd:nen
   iah te-ka-flat-owan’efl
   not NEG-NEUT.1TOWN-BE.BIG-STATIVE
   “It’s not a big town.”

The syntactic functions served by words in each lexical category are to some extent as might be expected. Morphological particles serve as demonstratives, numerals, question words, adverbials, conjunctions, discourse markers, etc. Morphological nouns function syntactically as arguments of clauses. Morphological verbs function syntactically as predicates. Verbs can also serve other functions. Since they contain pronominal prefixes specifying their core arguments, they can function as complete grammatical sentences in their own right, as can be seen in many of the preceding examples. Verbs can also be used as descriptive labels for entities.

(9) Morphological verbs as lexical nominals

a. iera’wistakarhathositha'
   ie-ra’wist-a-karhatho-hst-ha'
   INDEFAGENT-PANCAKE-EPENTHETIC-TURN.OVER-INTR-IMPERF
   “one flips pancakes with it” = “spatula”
Verbal nominals like these can function syntactically as arguments of clauses, without any overt marking of their status as nominals. A great many have been lexicalized as nominals, names of entities, much like English eggbeater. Their verbal origin can still be detected in their morphological behavior, however. Only formal noun stems can be incorporated into verbs. If verbal nominals are incorporated, they must be overtly nominalized with a suffix. The choice of nominalizer is a lexical matter. That for “table” is -'tsher-, and that for “car” is -'ht.

(10) wa'ke'serehtahn'tjon
wa'k-'sere-h't-a-hmnon-
FACT-1SG.AGENT-drag-NOMINALIZER-EPENTHEThIC-buy-PERF
“I bought a car.”

2. Lexicalization, headedness reversal, and decategorialization

The lexicalization of morphological verbs as names of entities has set the stage for a significant grammatical development. The head of an endocentric construction is usually taken to be that element which determines grammatical properties of the whole. The head of a noun-verb compound like -khw-onni “food-make” = “cook” is the verb root -onni “make”, since the resulting stem is a verb. The incorporated noun root serves as a modifier, describing a kind of preparation.

Certain stative verbs with incorporated nouns are evolving in an interesting direction. The verbs -iio “be good” and -aksen “be bad” can be seen in their basic predicative function in (11).

(11) Mohawk statives as predicates: Awenhrathen Deer, speaker
ótia'ke kanahskwtiio
ótia'ke ka-nahskw-iio
some NEUT.1-DOMESTIC.ANIMAL-Be.GOOD.STATIVE
ótia'ke ni' kanahskwáksen
ótia'ke ni' ka-nahskw-aks-en
some CONTRAST NEUT.1-DOMESTIC.ANIMAL-Be.BAD.STATIVE
“Some were good animals, some were bad animals.”
The same morphological construction, based on the same verb roots, shows a reversal of headedness in some contexts. The last word in (12) below shows the internal morphological structure of a verb, with the literal translation “it is a good thing”. It is used syntactically as a nominal, however, and its free translation is “good will”. The incorporated noun -rihw- “matter, affair, idea, thing” has become the head of the word, and the original verb -iio “be good” serves simply as a modifier “good”.

(12) Headedness reversal: statives as modifiers Warisose Karierithon, speaker
rati'hawi: ne raoneriáhsakon
rati-hawi:
MASC.PL.AGENCY-CARRY-IMPERF the MASC.PL.POSS-HEART-BE.IN-STATIVE
ne skén:nen’ tánon’ ne karihwi:io.
ne skén:ne’ tánon’ ne ka-rihw-íio
the peace and the NEUT.I-MATTER-BE.GOOD.STATIVE
“They had hearts full of peace and good will.”

As noted earlier, both verbs and nouns contain prefixes distinguishing gender and number. Those on verbs are pronominal prefixes that refer to the core arguments of the clause. They distinguish two grammatical relations based on semantic role. One paradigm of prefixes (Paradigm I) generally represents grammatical agents (ka-tákhe “it is running”), a second paradigm (Paradigm II) represents grammatical patients (ió-ta’s “it is sleeping”), and a third paradigm (transitive) represents combinations of agents and patients (taahshákóhsere’ “he was chasing them”). Arguments of inherent states are generally represented by Paradigm I prefixes (ka-kowá:nen “it is big”), while participants affected by resultant states are generally represented by Paradigm II prefixes (ió-tarihen “it is hot”). Perfects are classified grammatically as resultant states (ió:kon “it has eaten”).

The prefixes on nouns simply indicate gender and, for humans, number. The vast majority of nouns are neuter, because the majority of nouns refer to objects. There are two paradigms of noun prefixes, but unlike the pronominal prefixes on verbs they do not distinguish grammatical role. The choice of prefix paradigm is simply lexicalized with each noun: ka-nákta’ “bed” (neuter, Paradigm I), o-nó:wa’ “guitar” (neuter, Paradigm II). The choice is invariant for each noun, unrelated to the syntactic role of the noun in the clause. As can be seen in (13), the noun for “mouse” always appears with a Paradigm II prefix (o-), whether the coreferent pronominal prefix on the accompanying verb is from Paradigm I (agent ka-) or Paradigm II (patient io-).

(13) Noun prefix paradigm: independent of syntactic function
a. Katákhé’ ne otsinó:wen.
ka-takhe’ ne o-tsinow-en’
NEUT.I-TUN-IMPERF the NEUT.II-MOUSE-RELF.SUFFIX
“The mouse is running.”
b. I￥:ta$'s ne ￥tsin$wen.
    io-ta$'s ne ￥tsinow-en$'
    NEUT.II-sleep-IMPERF the NEUT.II-mouse-NOUN.SUFFIX
    "The mouse is sleeping."

The shapes of the pronominal prefixes on verbs and gender prefixes on nouns are similar but not identical. A comparison of the prefixes in the left and right columns in (14) below shows that many of the verb prefixes contain an initial glide that is absent from noun prefixes. (The letter $ represents the glide [y] before a vowel. Digraphs en and on represent nasalized vowels.)

<table>
<thead>
<tr>
<th>(14) Neuter prefixes on verbs and nouns</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Paradigm I</strong></td>
</tr>
<tr>
<td>AGENT</td>
</tr>
<tr>
<td>&quot;it is running&quot;</td>
</tr>
<tr>
<td>w-</td>
</tr>
<tr>
<td>&quot;it disappears&quot;</td>
</tr>
<tr>
<td>i-</td>
</tr>
<tr>
<td>i-onhe'</td>
</tr>
<tr>
<td>&quot;it is alive&quot;</td>
</tr>
<tr>
<td><strong>Paradigm II</strong></td>
</tr>
<tr>
<td>PATIENT</td>
</tr>
<tr>
<td>&quot;it is sleeping&quot;</td>
</tr>
<tr>
<td>iaw-</td>
</tr>
<tr>
<td>&quot;it is dead&quot;</td>
</tr>
</tbody>
</table>

When verbs contain incorporated nouns, the choice of pronominal prefix paradigm is normally governed by the verb root, since it is the head of the word. The verb -nor- "be precious", for example, requires a Paradigm I pronominal prefix, as can be seen in (15a). When the noun "silk" is incorporated, as in (15b), the Paradigm I prefix ka- remains unchanged. When the noun "wood" is incorporated as in (15c), the Paradigm I prefix ka- still remains unchanged, even though the noun "wood" by itself appears with the Paradigm II prefix (o-).

<table>
<thead>
<tr>
<th>(15) Prefix paradigm governed by verb</th>
</tr>
</thead>
<tbody>
<tr>
<td>ka-nor-on</td>
</tr>
<tr>
<td>NEUT.I-be.precious-STATIVE</td>
</tr>
<tr>
<td>&quot;It is precious, dear, expensive.&quot;</td>
</tr>
</tbody>
</table>
    ka-’nhehs-nor-on
    NEUT.I-silk-be.precious-STATIVE
    “It is expensive silk.”

    ka-ient-nor-on
    NEUT.I-wood-be.precious-STATIVE
    “It is expensive wood.”

In certain combinations, however, the prefix choice is now governed by the incorporated noun. These are combinations that have been lexicalized as nominals, like the term for “gold” in (16). Here the headedness relation has been reinterpreted: the incorporated noun is the head, and the incorporating verb a modifier.

(16) Prefix paradigm determined by incorporated noun

   ŧhwistanó:ron
   ŧ-hwist-nor-on
   NEUT.II-metal-be.precious-STATIVE
   “it is precious metal” > “precious metal” = “gold”

Though the verb -nor- “be precious” normally requires the Paradigm I prefix ka-, the prefix on “gold” matches that of the incorporated noun “metal” (II). Furthermore, the prefix shows the shape used with nominals (o-), rather than that used with verbs (io-). There is also a shift in referentiality. When verbs function as predicates or clauses, it is the pronominal prefix which is referential: “it is precious metal”. When they function as nominals, the whole construction is referential: “gold”.

The government of prefix choice by the incorporated noun is not an across-the-board structural change in all verbs serving as nominals. It is limited to a subset of lexical items containing only certain stative roots. These roots have the kinds of meanings typical of adjectives found in languages with limited adjective categories: good/bad, large/small, long/short, old/new, etc. (Dixon 1977 [1982]). The Mohawk root -as- “be new/fresh” is also part of the set. It normally appears with Paradigm I prefixes.

(17) Paradigm I verb: -as- “be new/fresh”

   wa:se’
   w-as-e’
   NEUT.I-be.new-STATIVE
   “it is new, fresh”

When the verb is reanalyzed as a modifier, the incorporated noun can govern prefix choice.
Apart from the fact that their prefix paradigm is governed by the incorporated noun rather than the verb root, these attributive adjective constructions still show the morphological structures and patterns of verbs with incorporated nouns. The incorporated element must either be a noun root or carry an overt nominalizing suffix.

(19) Overt nominalizer with -iio “good”

a. ka’serehtt’iio
   ka-’sere-ht-iio
   NEUT.1-drag-NOMINALIZER-be.nice.STATIVE
   “it’s a nice car”

b. watekhwahra’tshert’iio
   w-ate-khw-hra-’tsher-iio
   NEUT.1-MIDDLE-food-set.on-NOMINALIZER-be.nice.STATIVE
   “it’s a nice table”

(20) Overt nominalizer with -ase”“new”

a. ka’serehtasé’tsi
   ka-’sere-ht-as-e’-tsi
   NEUT.1-drag-NOMINALIZER-be.new-STATIVE-INTENSIFIER
   “new car”

b. atekhwahra:tsheres
   (zero)-ate-khw-hra-’tsher-es
   NEUT.1-MIDDLE-food-set.on-NOMINALIZER-be.long.STATIVE
   “long table”

Comparative constructions are formed with Mohawk adjectival verbs by means of the particle sénha “more”: sénha kakowá:nen “it is bigger” (also sénha kowá:nen).
The particle does not appear with nominals containing attributive adjectival suffixes, however, perhaps for structural reasons. The particle “more” would modify the word as a whole, rather than just a portion of it (a dependent qualifier), and the result would make little sense.

The kinds of situations that might prompt the reversal in headedness are easy to find. Sentences like that (21), from a conversation, are potentially structurally ambiguous.

(21) Structural ambiguity: Sha'tekenhatie' Phillips, speaker

Né: ki’ kheientéri,
ne: ki’ khe-ienter-i
it.is just 1SG/INDEF-KNOW-STATIVE
"Those are the ones I used to know,
Iako-itiohkwá:nen ... Oaks.
Iako-itiohkow-owan-en
INDEF.II-GROUP-LARGE-STATIVE (name)
they were a large group, the Oaks family/a large group, the Oaks family”.

The use of morphological verbs as syntactic nominals facilitated the reinterpretation of headedness in incorporating constructions. The lexicalization of such complex morphological structures as single units of expression dimmed speaker awareness of their internal structures.

Comparative evidence confirms the sequence of events. Cherokee, the sole member of the Southern branch of the Iroquoian family, does not show the same elaboration of noun incorporation as the Northern languages, nor the same reanalysis of certain stative verbs in certain contexts as attributive adjectives. An adjective category has begun to emerge in that language as well, but from a different diachronic source (Lindsey & Scancarelli 1985).

3. Further grammaticalization

Lexicalization has moved the grammar of the Northern Iroquoian languages ahead still one more step, but only with a few morphemes that recur frequently in lexicalized constructions. The stative verb -kowanen “be large”, has also evolved a step further into a derivational clitic =kó:wa. This new marker functions as an augmentative nominalizer, which can be added to any lexicalized nominal, of either nominal or verbal origin, to create new lexical nominals.
Similar developments have occurred with several other markers that originated as incorporating stative verbs, including a diminutive and set of locatives.

4. Conclusion

The processes of lexicalization and grammaticalization have often been compared in recent work on the factors that shape language. Both involve a cognitive process of routinization, the automation of frequently recurring sequences, but they differ in their output: the first creates lexical items (such as eggbeater or ohserá:se' “New Year’s Day”), while the second, as now more generally understood, creates grammar: grammatical markers and constructions (such as the oft-cited English “be going to” future or SVO word order). Neither process operates in isolation, apart from a linguistic and pragmatic context. Here we have seen an example of the potential role of the first (lexicalization) in providing an especially tight linguistic context for the second (grammaticalization). The routinization of recurring morpheme sequences as unitary lexical items resulted in a fading of their internal semantic and grammatical structure. This fading facilitated the reanalysis of the internal structure in terms of their external syntactic uses. The result was a new lexical category for certain stative verb roots as attributive adjectives. This process can still be seen to be constrained by lexicalization, occurring only with certain roots in certain lexical constructions. Finally, some lexicalized constructions containing these attributive adjectives provided the context for further grammaticalization. Some of the attributive roots are now evolving into derivational nominalizers.

Lexicalization is of course a gradient phenomenon. Morphological verbs used as syntactic nominals in Mohawk show varying degrees of lexicalization as nominals. Some, like atekhw:ra “table”, are so fully lexicalized as nominals that speakers no longer have conscious access to their verbal etymologies. Some, like kásere “car/it drags”, are robustly lexicalized as nominals, so that speakers typically provide an English noun as the first translation (“car”), but their literal verbal origins are still accessible. Some verbs are used spontaneously to designate entities, and the nonce forms may or may not subsequently become routinized with this function. And of course lexicalization does not necessarily stop at word boundaries. As we know from English, recurring sequences of words may become lexicalized as

<table>
<thead>
<tr>
<th>Augmentative nominalizer</th>
<th>takó:s</th>
<th>“cat”</th>
<th>takó:s=kó:wa</th>
<th>“wildcat, panther”</th>
</tr>
</thead>
<tbody>
<tr>
<td>d:niate’</td>
<td>“snake”</td>
<td>ohniare’=kó:wa</td>
<td>“serpent”</td>
<td></td>
</tr>
<tr>
<td>otsinó:wen</td>
<td>“mouse”</td>
<td>otsinowen’=kó:wa</td>
<td>“rat”</td>
<td></td>
</tr>
<tr>
<td>kahonwé:ia</td>
<td>“boat”</td>
<td>kahonweia’=kó:wa</td>
<td>“ship”</td>
<td></td>
</tr>
<tr>
<td>onon’ónsera’</td>
<td>“squash”</td>
<td>onon’ónsera’=kó:wa</td>
<td>“pumpkin”</td>
<td></td>
</tr>
<tr>
<td>kaniá:tará</td>
<td>“lake”</td>
<td>kaniatara’=kó:wa</td>
<td>“ocean”</td>
<td></td>
</tr>
</tbody>
</table>
idioms, though speakers typically retain access to their components longer than they may to morphemes within words, particularly in languages with extensive phonological fusion and little tradition of writing. But in all cases, lexicalization can dim the salience of internal grammatical and semantic structure, a development that can facilitate the reanalysis of the components of the construction and their functions.

Notes

* I am grateful to the Mohawk speakers from the communities of Kahnawake, Kanehsatake, Akwesasne, Thaientaneken, Wahta, and Ohsweken, who have generously shared their expertise. I especially appreciate insightful comments provided by Kanerahtenhawii Nicholas and Skawennati Montour of Kanehsatake, Kaia’titakhke Jacobs of Kahnawake, and Rokwaho Dan Thompson of Akwesasne.

1. Material cited here is presented in the practical orthography adopted by all six Mohawk communities. Symbols $t$ and $k$ represent plain stops (voiced before voiced segments); $ts$ is an alveolar affricate in Kahnawake, Kanehsatake, and Wahta, which corresponds to an alveopalatal affricate spelled $ts$ in Akwesasne, Thaientaneken, and Ohsweken; $s$ is a voiceless spirant; $n$, $r$, $w$, $i$ are resonants, with $i$ representing the glide [y] before vowels; $h$ is always pronounced as a distinct segment ($th = t+h$); and the apostrophe ' represents glottal stop. The vowel symbols $i$, $e$, $a$, $o$ have IPA values. Digraphs $en$ and $on$ represent nasal vowels: $en$ is a low, central vowel ( nasalized caret) and $on$ is a high, back vowel ( nasalized [u]). The colon : represents vowel length, the acute accent ' stress with high or rising tone, and the grave accent ` stress with falling tone.

References


Thomas, Elaine 1978. A Grammatical Description of the Engernni Language. Arlington, TX: University of Texas and Dallas, TX: Summer Institute of Linguistics.