A Nonsyntactic Template for Syntactic Noun Incorporation

Michael J. Houser and Maziar Toosarvandani
University of California, Berkeley
LSA Annual Meeting, January 7, 2006

1 Introduction

Baker (1996:25-30) observes that syntactic noun incorporation places the incorporated noun (IN) before the verb root, regardless of the language’s basic word order. In order to account for this, Baker posits a universal syntactic constraint on movement (a la Kayne 1994) that only permits left head-adjunction (1).

(1) If X and Y are X0 categories and X is adjoined to Y in the syntax, then X precedes Y in linear order (Baker 1996:29).

The same syntactic principle predicts that tense, aspect, and mood (TAM) morphology should follow the verb root. The verb raises successively through each functional projection in order to pick up TAM suffixes, undergoing head-adjunction according to the principle in (1):

(2) Incorporation then proceeds successive cyclically from the bottom of the tree to the top, adjoining the head of each phrase to the left of the next head, in accordance with [1]. This derives [N-V-Asp-Infl] as the unmarked morpheme order (Baker 1996:30).

Agreement morphology, in contrast, precedes the verb root, since it is introduced by head-adjunction to the Infl and Asp heads (Baker 1996:32). The verb, when it raises, adjoins to the head of the target phrase, ending up inside of agreement morphology.

At the end of the derivation, the TAM morphology is on the opposite end of the verb root as the IN and agreement morphology.

Data from Mapudungun leads Baker et al. (2005) to retract the claim that the IN universally precedes the verb.

(3) If NI in Mapudungun is syntactic, as we claim, this means that Baker (1996:§1.6) was wrong to say that head movement always adjoins the moved noun to the left of the verbal head (Baker et al. 2005:144 fn. 6).

Despite the importance of ordering issues in typology, the position of the IN and TAM and agreement morphology with respect to the verb root has not had a systematic cross-linguistic survey.

We conclude that the cross-linguistic frequencies are not stark enough to support Baker’s (1996) claims that universal principles of the syntax dictate the position of the IN, TAM, and agreement morphology with respect to the verb.

*We present the findings of a research group at the University of California, Berkeley, which also includes Gabriela Caballero, Teresa McFarland, Nicole Marcus, Anne Pycha, Suzanne Wilhite, and Johanna Nichols.
2 Definitions

2.1 Incorporation

⇒ We consider a noun to be incorporated if it structurally forms an inflectional unit with the verb stem or root. For consistency, when doing a large survey of grammars treating incorporation in varying degrees of detail and with varying theoretical assumptions, we used the position of the IN between parts of the inflected verbal complex as a necessary condition for incorporation.

A classic example of incorporation comes from Mohawk, where the IN appears between an agreement prefix and the verb stem.

(7) Mohawk (Iroquoian: U.S. and Canada)
ra-wira-muhwe'-s
he-baby-like-HAB
\[Baker\ (1997:279)\]

Verbal inflectional morphemes, which comprise part of the verbal complex, need not be affixal. In Yeli Dnye the verbal inflectional affixes cluster together in an auxiliary that precedes the verb root and is phonologically discrete and written as a separate word (8). The IN goes between the auxiliary and the verb stem, as shown in (9); we therefore consider this to comprise an instance of incorporation.

(8) Yeli Dnye (isolate: Melanesia)
D:a
pi.im.pst.1sg.S.cls
pˆeˆed
pct
\[Henderson\ (1995:16)\]

(9) Nmee-n:aa
yi.piâa
paapaa
cl.rem.1pl.S-mot
tree.log
pulling
\[Henderson\ (1995:27)\]

Our criteria rule in (10) from Car Nicobarese as an instance of incorporation. The subject agreement morpheme an doubles the independent subject and is therefore considered an agreement marker. Though in some respects it is word-like and is written as a separate word by Braine (1970), the object noun is inserted between the verb and this agreement marker.

(10) Car Nicobarese (possibly Austroasiatic: Nicobar Islands)\footnote{The “,” here marks a boundary type not clause juncture as in the orthography for English.}
ting/n
lotumk\ an, enuus
sent away Tatmak he James
\[Braine\ (1970:179)\]

This position is the obligatory one for objects, and so object noun incorporation in Car Nicobarese is obligatory. The situation is similar for Saweru (see Appendix C) and Onge (not in our sample; unclassified: Andaman Islands: Mark Donohue, p.c.) These examples are not called incorporation in the grammars, and could possibly also be analyzed as cliticization of the subject marker to the object but they certainly bear analysis as incorporation.

Our criteria rule out (11) from Samoan as incorporation, though it is called that in the grammar and has the incorporation-like properties of prosodic univerbation and detransitivization (i.e. pepe in (11) is not a syntactic direct object).

(11) Samoan (Austronesian, Polynesian: Samoa)
e
tausi pepe le teine
GENR nurse baby ART girl
\[Mosel \ and \ Hovdaugen\ (1992:255)\]

The Samoan process is syntactic and productive, but it is probably a better candidate for an ergative analog to differential object marking than incorporation.\footnote{Languages with overt case-marking of the direct object may mark some objects but not others on the basis of semantic or pragmatic criteria. The higher the object is on the scale, e.g. specificity, animacy, etc., the more likely it is to be marked overtly. Objects lower on the scale are more likely to be unmarked and form a prosodic constituent with the verb (Aissen 2003; Bossong 1985). This is the phenomenon called differential object marking.}

Our rigid definition of incorporation requires including Car Nicobarese and Saweru and excluding Samoan; the influence from this choice on our survey is statistically negligible.

2.2 Syntactic Incorporation

Following Baker et al. (2005), we defined syntactic incorporation as comprising Mithun’s (1984) Types II, III, and IV:

(13) Type

<table>
<thead>
<tr>
<th>Characteristic Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic, nonreferential;</td>
</tr>
<tr>
<td>Another NP takes on the grammatical function it vacates</td>
</tr>
<tr>
<td>Productive for discourse purposes, e.g. to background known information</td>
</tr>
<tr>
<td>An IN can be supplemented by more specific NP material external to the complex verb</td>
</tr>
</tbody>
</table>

Type II

In Tupinambá, incorporation of a noun allows another NP to become the direct object, e.g. the 3sg prefix in (14).

(14) Tupinambá (Tupi-Guarani: Brazil)
a-s-\text{ojá}-cy
I-him\ face-wash
\[Mithun\ (1984:457)\]
Type III

In Nahautl, the object kočillo is incorporated once it is old information.

(15) Nahautl (Uto-Aztecan: Mexico)
   where is  knife  I 1-it-want now
   ‘Where is the knife? I want it now.’
B: ya’ ki-kočillo-tete’ki punci.
   he (he)-it-knife-cut  bread
   ‘He cut the bread with it (the knife).’ [Mithun (1984:861)]

Type IV

In Mohawk, a generic IN restricts the verb while a verb-external NP provides the specific object reference.

(16) Mohawk
   several so.it.year.numbers so it goes eight of them bullhead be-fish-bought this my.father
   ‘Several years ago, my father bought eight bullheads.’ [Mithun (1984:870)]

2.3 Productive Incorporation

Syntactic incorporation is virtually always productive (p < 0.005), while nonsyntactic incorporation can be either productive or nonproductive.

(17)    | Productive | Unproductive | Total |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntactic</td>
<td>25</td>
<td>1</td>
<td>26</td>
</tr>
<tr>
<td>Nonsyntactic</td>
<td>4</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>29</td>
<td>5</td>
<td>34</td>
</tr>
</tbody>
</table>

⇒ We define productive incorporation as incorporation that can involve any member of a non-closed class of nouns.

Thus, Kutenai, which allows incorporation of a closed class of nouns, namely ‘hand’, ‘foot’, ‘mouth’, ‘body’, ‘finger’, ‘heat/fire’, and ‘blade’ into any verb, is not considered productive.

(18) Kutenai (isolate: Canada)
mat’-hiy-ni
   dirty-hand→IND
   ‘He dirtied his hands’ [Morgan (1991:345)]

In contrast, Halkomelem, in which body parts—a large if not open class of nouns—can incorporate into the verb, counts as productive.

(19) Halkomelem (Salishan: Canada)
lak*’-xén
   get.broken-leg
   ‘break a leg’ [Suttles (2004:307)]

(20) t’eq*-elwas-t
   punch-side-trans
   ‘punch him on the side’ [Suttles (2004:307)]

It is possible for the class of verbs that allow incorporation to also be restricted but we consider these cases to be cases of productive incorporation. In Tümpisa Shoshone, the class of possible verb stems is limited to five, but these allow any noun to incorporate into them.

(21) Tümpisa Shoshone (Uto-Aztecan: U.S.)
númmú so’oppih pusish pungkupaimmuñpiñútú
   we:exc many burro pet-have-HAB-PST
   ‘we used to have many burro pets’ [Dayley (1989:91)]

Similarly, for Kayardild, two verbs, -marutha ‘put’ and -barrwaaja ‘block’ allow incorporation of an oblique noun.

(22) Kayardild (Tangkic: Australia)
kabilji-marutha
   grave-put
   ‘lay in grave’ [Evans (1995:294)]

(23) kalarr-barrwaaja
   open.space-block
   ‘prevent from coming out into the open’ [Evans (1995:295)]

Note in both these cases, the members of the closed class of verbs have an independent existence and combine in a semantically transparent manner.

3 Methods and Data

Our sample consisted of 66 languages. 46 incorporation patterns were found, of which 28 were syntactic and 32 productive. (See attached map.)

3.1 Position of the IN

3.1.1 Preverbal IN is preferred overall

Incorporation, syntactic or nonsyntactic, exhibits a general tendency for preverbal IN position (n.s.)

(24)    | Pre | Post |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntactic</td>
<td>21</td>
<td>7</td>
</tr>
<tr>
<td>Nonsyntactic</td>
<td>8</td>
<td>2</td>
</tr>
</tbody>
</table>
3.1.2 Productive incorporation correlates with preverbal IN

Productive incorporation is sometimes postverbal but in our sample unproductive incorporation never is \((p < 0.05)\).

\[(25)\]

<table>
<thead>
<tr>
<th></th>
<th>Pre</th>
<th>Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>Productive</td>
<td>21</td>
<td>11</td>
</tr>
<tr>
<td>Unproductive</td>
<td>7</td>
<td>0</td>
</tr>
</tbody>
</table>

3.1.3 Position of IN correlates with word order in productive incorporation

For productive incorporation, position of the IN is sensitive to language’s word order \((p < 0.025)\).

\[(26)\]

<table>
<thead>
<tr>
<th></th>
<th>Pre</th>
<th>Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>VO</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>OV</td>
<td>8</td>
<td>1</td>
</tr>
</tbody>
</table>

3.2 Position of TAM and Agreement Morphology

⇒ There is a slight tendency in languages with incorporation for TAM and agreement morphology to occur on opposite ends of the verbs. The opposite tendency occurs in languages without incorporation. (Half points = split patterns. Bold = predicted largest by Baker.)

\[(27)\]

<table>
<thead>
<tr>
<th></th>
<th>Same</th>
<th>Different</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorporation</td>
<td>13.5</td>
<td>15.5</td>
</tr>
<tr>
<td>No incorporation</td>
<td>8.5</td>
<td>4.5</td>
</tr>
</tbody>
</table>

This skewing is not statistically significant.⇒ The position of the IN does not correlate with the position of TAM marking. For syntactic incorporation, the correlation merely approaches significance \((p > 0.05)\).

\[(28)\]

<table>
<thead>
<tr>
<th></th>
<th>TAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre</td>
<td>5</td>
</tr>
<tr>
<td>Post</td>
<td>0</td>
</tr>
</tbody>
</table>

For productive incorporation, no correlation is observed at all.⇒ But, there is a significant tendency for languages with incorporation to have preposed TAM and/or agreement morphology \((p < 0.05)\). There is no difference between syntactic and non-syntactic incorporation in this regard.

\[(29)\]

<table>
<thead>
<tr>
<th>TAM AND/OR AGREEMENT</th>
<th>Pre</th>
<th>Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorporation</td>
<td>21</td>
<td>8</td>
</tr>
<tr>
<td>No incorporation</td>
<td>5</td>
<td>8</td>
</tr>
</tbody>
</table>

This result is most likely artifactual. Our strict definition of incorporation requires some prefixal inflection given that INs are more likely to be preposed.

4 Analysis & Conclusion

⇒ Baker’s earlier claim for syntactic noun incorporation that the IN universally precedes the verbal root is not founded. While preverbal IN is preferred, it is not categorical. Nor does a statistically significant correlation exist.

⇒ Interestingly, conformity to Baker’s generalization is strongest in unproductive and fossilized contexts and counterexamples occur precisely where syntactic derivation should be most transparently reflected.

⇒ Something other than universal constraints on movement in the syntactic derivation must be responsible for this preference for preverbal IN order.

- We suggest that the output of syntactic incorporation is probabilistically attracted to a universally favored ordering principle that shapes other kinds of noun-verb combinations as well and is independent of syntactic principles.
- The theoretical status of templates in typology and formal theory is not settled, but preverbal IN may be an example of one. In diachronic terms, older, unproductive incorporation universally conforms to the template, while newer and/or non-affixal incorporation sometimes follows the language’s word order rather than the template. That is, the longer incorporation has been in the language the greater the chance that it will conform to the template.

⇒ The position of TAM and agreement morphemes shows a modest preference for what Baker predicts, though it is not strong enough to support his account for it, which again derives from universal principles of the syntax.

Appendix A: Compounding

Survey (in progress) of agent/instrument and event nominalizations such as \textit{skyscraper} and \textit{witch hunt}, which contain both a noun and a verb root. There are two competing forces: (a) the proposed template which favors N + V order; (b) the language’s own word order, with OV favoring N + V and VO favoring V + N.

Expected frequencies (if template and word order compete) are given in (30). The cells in boldface represent patterns in which the language’s basic word order does not dictate the order in compounds and some other force favors N-V order. The cells that are not bolded represent patterns that are fully accounted for by the language’s basic word order.

\[(30)\]

<table>
<thead>
<tr>
<th></th>
<th>N-V</th>
<th>V-N</th>
</tr>
</thead>
<tbody>
<tr>
<td>OV</td>
<td>many</td>
<td>none</td>
</tr>
<tr>
<td>VO</td>
<td>some</td>
<td>none</td>
</tr>
</tbody>
</table>

⇒ The actual findings are given in (31). (* = majority or productive type, for languages that have both orders.)
Appendix B: Head- vs. dependent-marking and incorporation

Baker (1996) argues that head-marking is a necessary requisite for syntactic incorporation.

(35) ... languages in which the MVC [Morphological Visibility Condition: the principal driving NI] holds can never get by with only noun incorporation; rather they must have rich subject and object agreement paradigms as well... Thus, languages with productive NI must be of the head-marking type (Baker 1996:282–283).

⇒ Languages with syntactic incorporation are head-marking nearly twice as often as expected, and almost never dependent-marking.

<table>
<thead>
<tr>
<th>Incorporation</th>
<th>Head</th>
<th>Double</th>
<th>Dependent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Languages</td>
<td>17</td>
<td>7</td>
<td>1</td>
</tr>
</tbody>
</table>

(Implication counts come from our sample of languages with syntactic incorporation. No incorporation counts come from the Autotyp worldwide sample minus our sample of incorporating languages.)

However, it is difficult to determine whether this is a grammatical preference or a geographical coincidence. Incorporation and head-marking are both common in the Pacific Rim microarea. Head-marking is moderately well-attested and syntactic incorporation virtually nonexistent in Africa, a distribution that supports the geographical explanation.

Appendix C

Languages with noun incorporation in our sample with examples from some of them.

Ainu (isolate: Japan and Sakhalin Islands)

(37) a. asir cise ci-kar kor... new house 1PL.EX-make and ‘We made a new house and...’

(38) b. ney ta cise-kar-as there at house-make-1PL.EX ‘We made a house there’ [Shibatani (1990:61)]

Alamblak (Sepik Hill: New Guinea)

(39) a. hak#ku-nas-sininh-sa’ golden otter APP-war-begin ‘The war started because of the golden sea otter’ [Shibatani (1990:63)]

Caddo (Caddoan: southern U.S. plains)

(40) a. shi ku-nas-sininh-sa’ golden otter APP-war-begin ‘The war started because of the golden sea otter’ [Shibatani (1990:63)]

Compounds like breakwater, bilipoy, scarecrow, and turncoat are much less frequent, stylistically marked, and probably unproductive.
(40) ha#’ahat kn#ba-t-k’as-ya’ah
ADJ-good NEG-1DAT IRR-APP-leg-be
‘my wheel isn’t any good’ [Melnar (2004:48)]

(41) ‘ich’ah-na=’ni’-chah kassi
eye-DIST-buy-INTENT bead
‘he’s going to buy beads’ [Melnar (2004:174)]

(42) El ca¸cador va trenzar la cama a l’ocell
the hunter PST break-leg to the bird
‘The hunter broke the bird’s leg’

(43) El ca¸cador va cana-trenc-ar l’ocell
the hunter PST leg-break-leg to the bird
‘The hunter broke the bird’s leg(s)’ [Adelman (2002:5) citing Gracia and Fullana (1999:249)]

(44) taN-am@nan ya-qora-nm-at-len
INTS-alone personal.name.3SG.ABS PERF-reindeer-kill-VIB-3SG
‘C@kwaNaqaj all by himself slaughtered reindeer.’ [Dunn (1999:222)]

(45) a. -ihcikw¯an-
‘knee’

b. kask ihcikw¯anhew
‘he breaks his knee by shot’ [Wolfart (1973:67)]

(46) (Che) ai-po-pete la-mita
1 AC-hand-slap DEF-child
‘I slapped the child in the hand’ [Velazquez-Castillo (1996:99)]

Halkomelem (Salish: Canada)

(47) soko-hu
goodness-do [Haiman (1980:117)]

Kayardild (Tangic: Australia)

(48) Barri-ganj-ngune-ng
3 A/3P meat-eat-TNS
‘They ate the meat.’ [Evans (2003:330)]

(49) Bannuru a-godj-bom
magpie gose 1/3-head-hit-TNS
‘They hit the magpie goose in the head.’ [Evans (2003:330)]

Murrinh-Patha (Southern Daly: Australia)

(50) dem-NG-1-IO-back-heat
‘My back feels hot; something heats my back’

(51) thunku fire dem-NG-1-IO-back-heat
‘The fire makes my back feel hot’ [Walsh (1987); retranscription]

Ngandi (Gunwingguan: Australia)

(52) ngagu-jundu-geyk-dh-i
1SG>3SG stone-throw-? TNS
‘I threw a stone’ [Heath (1978:118-19); retranscription]

Nisgha (Tsimshian: western Canada)

(53) c´ap-yi:q-’a
canoe-travel in-TEL
‘he traveled in a canoe’ [Nakayama (1997:48)]

(54) mahti-q-(ˇc)i:l [L]-ˇsi-ˇcip
house-BFR make-PERF-BEN
‘he built her a house’ [Davidson (2002:188)]

Oneida (Iroquoian: northeastern U.S.)

(55) la-ahy-k-s s’d yes
PRO-fruit-eat-SER blackberries
‘he’s (fruit-)eating blackberries’ [Abbott (2000:63)]
Pipil (Uto-Aztecan: El Salvador)

Samoan (Austronesian, Polynesian: Samoan)

(56) e tausi pepe le teine
GENR nurse baby ART girl
‘the girl is a babysitter’

(57) e tausi pepe e le teine
GENR nurse baby ERG ART girl
‘the girl takes care of the babies’

[Saweru (West Papuan: New Guinea)]

(58) [auna ma wo-mo] mo=[V]o[rama] a-bai]
woman 3sgF.ERG 3sgF.NOM=man 3sg.M-hit
‘The woman hit the man’

[Saweru (West Papuan: New Guinea)]

(59) gl-bpe-te-ji
cut-buffalo-TNS-3PLS
‘They are slaughtering the buffalo’

[Slave (Athabaskan: Canada)]

(60) yede ti-shut-pe-ban
that 1sg.A-shirt-make-past
‘I made that shirt’

[Sora (Austroasiatic: Munda, India)]

(61) hliawrade
lady 0-seuan-mu-ban
A.A-man-see-past
‘the lady saw the man’

[Southern Tiwa (Tanoan: U.S.)]

(62) a-diru-k’ar-hi
2sg.A-chicken-eat-fut
‘You will eat the chicken’

[Totonac (Totonac-Tepelua: Mexico)]

(63) i-k’uru-k’euwe-m
B-dipper-old-pres
‘the dipper is old’

[Washo (isolate: western U.S.)]

(64) ji-mon-taqkina
he-me-steal
‘He stole it from me’

[Osbourne (1974:47)]

(65) ji-mon-kaui-qa
he-me-hand-grab
‘He grabbed me by the hand’

[Osbourne (1974:47)]

Tukang Besi (Austronesian, Malayo-Polynesian: Indonesia)

(66) No-sai kui-kui-mo
3r-make cakes-PERF
‘S/he has made cakes’

[Donohue (1999: 168)]

Tümpisa Shoshone (Uto-Aztecan: western U.S.)

Warembori (isolate: New Guinea)

Wawi’ (Chapakuran: Brazil)

(67) Hu capam’ in rain pacun!
blew cornbread completely 2sg.RP-3N stone
‘Turn the stones into cornbread!’

[Everett and Kern (1997:386)]

Yel Duyu (isolate: Melanesia)

Yimas (Lower Sepik: Papua New Guinea)

(70) ura-mpu-na-akpi-api-n
fire:O-3PL def back:ncm:sing-put.in-pres
‘They are putting (their) backs to the fire.’ (to warm themselves)

[Foley (1991:320)]

Yucatec (Mayan: Mexico)

[Abbreviations]

The abbreviations we use are as follows: 1 = 1st person; 2 = 2nd person; 3 = 3rd person; 3/3P, 1/3 = subject-object prefix complexes; A = agent; ABS = absolutive; ADJ = adjective; APP = applicative.
ART = article; BEN = benefactive; BFR = buffer consonant; CI = continuous indicative; CLS = close to speaker; DAT = dative; DEF = definite; ENG = ergative; FUT = future; GENR = Samoan tense category; HAB = habitual tense; IM = immediate; IMPF = imperfective; IND = indicative; INTS = intensifier; IRR = irrealis; IV = a gender; M = masculine; MOT = motion; N = neuter; NCM = noun class (gender) marker; NEG = negation; O = object; PI = punctiliar indicative; PCT = punctiliar; PRRF = perfective; PRES = present; PROG = progressive; PST = past tense; REM = remote past tense; RF = realis future; RP/R = reals past/present; S = subject; TEL = telic; TNS = a tense; VB = verbal derivational suffix.

References