

Nonsyntactic ordering effects in noun incorporation

Gabriela Caballero, Michael J. Houser, Nicole Marcus, Teresa McFarland, Anne Pycha,
Maziar Toosarvandani, Suzanne Wilhite, Johanna Nichols

Abstract

Despite the importance of ordering phenomena in typology and the visibility of Baker's analysis (1988, 1996) of noun incorporation in generative syntax, his prediction (1996:25-30) that in syntactic incorporation the incorporated noun will always precede the verb root has yet to be tested typologically. Here we fill this gap and survey the known cases of object noun incorporation. The predicted order proves to be strongly preferred cross-linguistically and warrants recognition as a strong statistical universal. However, it is strongest in unproductive and fossilized contexts, the opposite of what is expected if the position of the incorporated noun is determined solely by principles of syntactic movement. The universal must therefore be non-syntactic, perhaps morphological, in nature and appears to involve a preferred position for heads and/or for noun and verb roots. The same principle also shapes other noun-verb combinations in addition to noun incorporation.

Keywords: noun incorporation, compound, word order, morpheme order

1. Introduction

Despite the importance of ordering of elements in typology (Greenberg 1963 and much subsequent work, e.g. Dryer 1992, 1997, including in morphology, e.g. Bybee et al. 1990, Enrique-Arias 2002, Bauer 2001) and the perennial interest of noun incorporation to morphosyntactic theory (e.g. Mithun 1984, Sadock 1980, 1986, Baker 1988, Rosen 1989), the relative positions of the object and the verb in object noun incorporation (NI) have not been surveyed cross-linguistically.¹ This is the more striking in view of the visibility of Baker's analysis (1988, 1996) of noun incorporation (NI) in generative syntactic theory. Following work by Kayne (1994), Baker (1996:29) proposes that NI is the result of head movement of the noun into the verb and is restricted by a universal constraint that permits only left head adjunction, and that "If X and Y are X₀ categories and X is adjoined to Y in the syntax, then X precedes Y in linear order". Therefore the incorporated noun (IN) will universally precede the verb root, regardless of the basic word order in the language.

Not all elements of Baker's proposal have proven equally robust. Lexicalist and other non-syntactic analyses of incorporation continue to have currency in the literature (e.g. Rosen 1989, Spencer 1995, Gerdts 1998, van Geenhoven 1998, Malouf 1999, Runner and Aranovich 2003). Incorporated nouns are not always preverbal; Baker 1996:32 cites (1) from Sora (Austroasiatic: Munda; India), and Baker et al. 2005 explicitly retract Baker's

¹ [Acknowledgments to be added.] An earlier version of this paper was presented as Houser and Toosarvandani 2006.

earlier claim of categorically universal preverbal position for the incorporated noun on the evidence of Mapudungun (Araucanian; Chile), e.g. (2).²

- (1) Jom-**bɔŋ**-tɛ-n-ji pɔ?
 eat-buffalo-NONPAST-3SINTR-3PLS Q
 'Will they eat the **buffalo**?' or 'Do they eat **buffalo**?' (Baker 1996:23)

² Abbreviations: 1 = first person, 2 = second person, 3 = third person, 4 = fourth person; 3/3P, 1/3 subject-object prefix complexes; A = agent, ABS = absolutive, ACC = accusative, ADJ = adjective; AFF = affix, APP = applicative, ART = article, ATTR = attributive, BEN = benefactive, BFR = buffer consonant, CI = continuous indicative, CLS = close to speaker, CONJ = conjugation marker, DAT = dative, DECL = declarative particle, DEF = definite, DETR = detransitivizer, DIST = distributive, EP = epenthetic vowel, ERG = ergative, F = feminine, EX = exclusive, FUT = future, GENR = a Samoan tense category, HAB = habitual tense, IM = immediate, IMP = imperative, IMPF = imperfective, INCL = inclusive, IND = indicative, INSTR = instrument, INTENT = intentive, INTR = intransitive, INTS = intensifier, IO = individual verification, observational orientation, IRR = irrealis, IV = a gender class, LOC = locative, M = masculine, MOT = motion, N = neuter, NCM = noun class (gender) marker, NEG = negation, NZ = nominalizer, O = object, PI = punctilinear indicative, PASS = passive, PCT = punctiliar, PERF = perfective, PL = plural, PRES = present, PRO = pronominal prefix or (Samoan) pronominal anaphor, PROG = progressive, PST = past tense, Q = question particle, R = realis, REF = referential stem, REFL = reflexive, REM = remote past tense, RF = realis future, RP/P = realis past/present, S = subject, SER = serial suffix, SG = singular, T/E = tense/evidential, TEL = telic, TNS = tense, VB = verbal derivational suffix.

(2) Ñi chao kintu-**waka**-le-y

my father seek-**cow**-PROG-IND.3sgS

'My father is looking for the cows'

(Baker et al. 2005:139)

Baker is the only one to have made observations about the position of the incorporated noun, and we test that observation here. Following usual typological practice, we will accept a statistically significant cross-linguistic preference as a universal whether or not it is categorical. We conducted a worldwide survey to seek answers to the following questions:

- Is the incorporated noun usually placed before the verb root regardless of the language's basic order of verb and object? If so, is the frequency of this ordering sufficient to warrant regarding it as a universal preference?

- Is there any correlation between the type of NI (as defined below) and the position of the IN? In particular, is there support for Baker's claim that specifically syntactic incorporation (defined in §2.3 below) produces preverbal incorporated nouns?

We will argue that preverbal incorporation is a robust (though not exceptionless) cross-linguistic tendency and we therefore retain Baker's observation as a (statistical) universal. It is, however, most strongly in evidence in unproductive incorporation, i.e. in the context where syntactic considerations should be least relevant to morpheme order.

2. Definitions

2.1. Noun incorporation

For consistency in doing a large survey of grammars that treat NI in varying degrees of detail and with various theoretical assumptions, we consider a noun to be incorporated if it forms a single morphological unit with the verb stem or root. That is, a noun was counted as incorporated if it occurred between parts of the inflected verbal complex. (3) is a classic example of incorporation from Mohawk (Iroquoian: U.S. and Canada) in which the incorporated noun *wir* 'baby' appears between an agreement prefix and the verb stem *nuhwe* 'like'.

(3) ra-wir-a-nuhwe'-s

he-baby-EP-like-HAB

'He likes babies'

(Baker 1997:279; M. Baker, p.c.)

To satisfy our definition for incorporation, verbal inflectional morphemes do not have to be affixal. (4) is an example from Yeli Dnye (isolate: Melanesia), where the verbal inflectional morphemes cluster together in a preverbal element that is phonologically discrete and written as a separate word, yet is not a syntactic word. In Yeli Dnye, the

incorporated noun is positioned between the preverb and the verb stem, as shown in (4), a valid instance of incorporation by our definition.³

- (4) D:a pêêd
 P1.IM.PST.1SG.S.CLS pull.PCT
 'I pulled it (a shark) in' (Henderson 1995:16)

- (5) Nmee-n:aa yi.pââ paapaa
 CI.REM.1PL.S-MOT tree.log pulling
 'We were pulling logs' (Henderson 1995:27)

Another pattern involving non-affixal inflectional morphology that satisfies our criteria for NI comes from Car Nicobarese (Austroasiatic: Nicobar Islands), exemplified in (6). The morpheme *an* doubles the independent subject and is therefore an agreement marker. Though it is phonologically word-like and written as a separate word in the grammar, the object noun is inserted between the verb and the agreement marker and is therefore a valid incorporated noun.⁴

³ Contrast (8) with German examples such as *Hans hat einen Brief geschrieben* 'Hans wrote a letter' (lit. 'Hans has a letter written'). Here the auxiliary verb *hat* 'has' is a syntactic word (rather than an isolating inflectional formative), so German is not counted as having incorporation.

⁴ The comma in (6) is not a punctuation mark but transcribes a boundary type. It should not be taken to suggest (as the English punctuation comma would) that the last word is a duplication, afterthought, or the like.

- (6) tí:nɲén tətɲák an, ce:ms
 sent.away Tətɲak he James
 'James sent Tətɲak away' (Braine 1970:179)

This position is apparently obligatory for objects;⁵ that is, NI in Car Nicobarese is obligatory. The situation is similar for Saweru (see Appendix A) and Onge (not in our sample; unclassified language of the Andaman Islands; Mark Donohue, p.c.). These examples are not called incorporation in the respective grammars and could possibly also be analyzed as cliticization of the subject marker to the object. That is, they do not necessarily require analysis as incorporation, but they certainly bear such an analysis, and our criteria classify them as obligatory NI.

A less clearcut example of incorporation comes from Polynesian languages, represented by Samoan in our sample. (7) illustrates what has regularly been called incorporation in the Polynesianist literature (e.g. Chung 1978:183-189, Mosel & Hovdhaugen 1992) and also by Mithun (1984:850): the object is non-specific, verb and object are prosodically univerted, and the subject is absolutive, showing that the verb is intransitive; contrast unincorporated (8) with definite article on the object and ergative case on the subject. (9-10) show the same options with a pronoun subject; here the word order differs, with verb and unincorporated object separated by the pronoun subject in (10).

⁵ More precisely, for one word of the object NP; Braine 1971:248 indicates that either the object noun or its modifying adjective can be incorporated, and the other remains unincorporated.

Examples with more than one modifier are unfortunately not given.

(7) E tausi pepe le teine
 TNS care baby ART girl
 'The girl is a babysitter'

(8) E tausi pepe e le teine
 TNS care baby ERG ART girl
 'The girl takes care of the babies'

(Mosel & Hovdhaugen 1992:255)

(9) E tausi-pepe 'oia
 TNS care-baby he
 'He takes care of babies'

(10) E tausi e ia pepe
 TNS care ERG he baby
 'He takes care of the babies'

(Chung 1978:183)⁶

Examples like (7) and (9) might better be classified as *stripping* (Miner 1986) or *quasi-incorporation* (Dahl 2004:216-219) or *pseudo-incorporation* (Massam 2001): the object is non-specific, immediately adjacent to the verb, and often shorn of some inflection. They might also be taken as an ergative variety of *differential object marking* (Aissen 2003,

⁶ Chung writes the hyphen in (9); Mosel & Hovdhaugen do not use a hyphen, but mention (p. 392) that verb and incorporated object are often written as one word by Samoans.

Bossong 1985), in which objects lower on a scale of specificity, animacy, and/or similar factors are likely to lack overt object marking and/or form a prosodic constituent with the verb. In at least some Polynesian languages, particles that are normally directly postverbal follow the entire verb+noun incorporating sequence, e.g. Samoan (Chung 1978:184):

(11) Po 'o aafea e tausi **ai** e ia tama?

Q PRED when TNS care **Pro** ERG he child

When does he take care of children?

(12) Po 'o aafea e tausi-tama **ai** 'oia?

TNS care-child **Pro** he

When does he babysit?

The clitic particle, boldface in (11-12), is a pronominal anaphor used when an oblique nominal is missing from its usual position due to questioning, relativization, or the like (Chung 1978:21, Chapin 1974). With *tama* 'child' flanked by the tensed verb and this clitic, (12) fits (at least barely) our definition of incorporation.

We were prepared to accept a noun at the edge of the inflected verb as incorporated if it displayed clear word-internal sandhi, but found no such cases. In Nuuchahnulth, word order and absence of a dummy IN make it clear that a noun at the left edge of the verb is incorporated (see Appendix A).

This definition makes it possible to capture all possible examples of incorporation without mixing in phenomena such as stripping and DOM. To survey a cross-

linguistically rare phenomenon like incorporation, capturing all possible examples but no extraneous ones is essential. The statistical impact of including or excluding Samoan, Car Nicobarese, and Saweru is small; it is tracked in footnotes below in §3.

In some languages the incorporated noun differs in form from its unincorporated counterpart. In Yana (isolate: California), nouns with monosyllabic stems have a final *-na*, which is not present on the incorporated form; if the incorporated noun stem ends in a short vowel other than *i*, *-i* is added; and initial *b* or *d* of an incorporated noun often undergoes lenition to appear as *w* or *r* (Sapir 1911:268). Thus the incorporated form of *bana* 'deer' is *wai*; that of *xana* 'water' is *xai*; and that of *auna* 'fire' is *au*. Many nouns in Nahuatl, as in many other Uto-Aztecan languages, in their citation forms have what is called an absolutive suffix, which appears whenever the noun does not have some other suffix; these absolutive suffixes do not appear on incorporated nouns. Differences between independent and incorporated forms of nouns can be greater and less automatic, however. In languages of the Salishan and Wakashan families (represented by Halkomelem, Nuuchahnulth, and Kwak'wala in our sample), the relationship between the two can be one of allomorphy or outright suppletion. For instance, Halkomelem has the lexical suffix *'a:ləs* corresponding to the independent noun *qál:əm* 'eye' (Galloway 1993:203, 518; see Gerdts 2003). The incorporated forms are usually called *lexical affixes* in descriptions of these languages, but Gerdts 2003 shows that they are morphosyntactic arguments and not distinct from incorporation, and we classify them as incorporation. This means that suppletion of the

noun stem on incorporation is no obstacle to our classification of the construction as incorporation if it meets the other criteria.⁷

2.2. Types of incorporation

For convenience in classifying incorporation patterns into syntactic vs. nonsyntactic, we sometimes refer to the widely used four types of incorporation identified by Mithun (1984) and shown in Table 1 and illustrated just below.

| Type | Characteristic properties |
|------------------------------------|--|
| I "Lexical compounding" | Incorporated noun is generic, nonreferential; N+V is conventional, institutionalized activity |
| II "Manipulation of case" | IN loses argument status; another NP takes on the grammatical function it vacates |
| III "Manipulation of discourse" | NI used productively for discourse, e.g. to background known information |
| IV "Classificatory NI" | IN can be supplemented by more specific NP material external to the complex verb |

Table 1: Mithun's (1984) four types of NI

⁷ In Nunggubuyu (Gunwinyguan, Australia; Heath 1984:463-468), upon incorporation and similar compounding many nouns supplete or display phonological and/or semantic irregularities while many others do not, further supporting our contention that suppletive and non-suppletive IN's are the same kind of entity.

2.2.1. Type I

In Nisgha (Tsimshianic; western Canada), the incorporated noun is referentially nonspecific and the verb describes an institutionalized activity, as shown in (13).

- (13) qùł-hón n'ì'y
 gut.something-fish me
 'I gutted fish.' (Tarpent 1987:792)

2.2.2. Type II

In Tupinambá (Tupi-Guarani: Brazil), incorporation of a noun allows another NP to become the direct object. In (14), 'face' is incorporated; the direct object is signaled by pronominal agreement on the verb.

- (14) a-s-oβá-éy
 I-him-face-wash
 'I face-washed him.' (Mithun 1984:857)

2.2.3. Type III

In Nahuatl (Uto-Aztecan: Mexico), a direct object is incorporated once it becomes old information. In (15) the noun *kočillo* 'knife' is introduced, but not incorporated. In (16), it is old information so it is incorporated.

(15) Kanke eltok kočillo? Na' ni-'-neki amanci.

where is knife I I-it-want now

'Where is the knife? I want it now.'

(16) ya' ki- kočillo -tete'ki panci

he (he)it-knife-cut bread

'He cut the bread with it (the knife).' (Mithun 1984:861)

2.2.4. Type IV

In Mohawk, an incorporated noun is generic and serves to restrict the meaning of the verb while an external NP provides the specific object reference, as in (17).

(17) Tokha niyohserá:ke tsi nahe' sha'té:ku nikúti: rabahbót wahu-tsy-ahnínu
several so.it.year.numbers so it.goes eight of.them bullhead he-fish-bought

ki rake'niha.

this my.father

'Several years ago, my father bought eight bullheads.' (Mithun 1984:870)

2.2.5. Compounding vs. classificatory incorporation

Rosen 1989 distinguishes compounding vs. classificatory noun incorporation (see also Gerdts 1998, Runner and Aranovich 2003). In compounding incorporation the verb becomes intransitive (as though the incorporated noun were not in the argument structure),

and the incorporated noun cannot have external modifiers; in classificatory incorporation the verb remains transitive and the incorporated noun can have external modifiers and the like (an example is (17) above; also Caddo in Appendix A). Classificatory incorporation generally falls into Mithun's Types II-IV (IV is always classificatory), and is generally syntactic in Baker's terms. It is almost always productive as defined below (Catalan is exceptional). Appendix B shows the compounding vs. classificatory type where identified in the literature and, where possible, additional cases identified by us.

2.3. Syntactic incorporation

Noun incorporation has been variously analyzed in the literature as a morphological process and as a syntactic process. In morphological approaches (e.g. Mithun 1984), NI is viewed as the result of presyntactic or lexical compounding of a noun and a verbal stem. Productive and transparent though NI often is, under this view it is nonetheless a word formation process contributing to a finite lexicon.

Later work within this perspective proposes that NI is a process that operates at the level of argument structure (Rosen 1989; Spencer 1995). Spencer's analysis of Chukchi (Chukchi-Kamchatkan: Russia) argues that Chukchi NI is a morphological N-V verb compound that is SEMANTICALLY equivalent to a transitive construction, where the verb's argument structure is morphologically saturated: NI is a morphological operation over argument structures.

Syntactic approaches, in contrast, posit a syntactic operation that is responsible for moving a noun into the verb of which it is an argument to form a single word. Sadock's (1980, 1986) defense of the syntactic nature of NI depends on evidence that the

incorporated noun interacts with the external syntax of the sentence in ways that require the noun to be represented at some level as independent of the incorporating verb. These include external modification or possession of the incorporated noun and the ability of the incorporated noun to serve as a discourse referent. In West Greenlandic (Eskimoan: Greenland), for example, an incorporated noun introduces a discourse referent that can later be picked out, in this language by person/number suffixes on the verb. In (18), the incorporated noun of the (a) sentence introduces 'airplane' as a discourse referent that the 3sg. and 4sg. suffixes in the (b) sentence refer to.

- (18) a. Suulut timmisartuliorpoq
 Søren.abs airpane-make-INDIC-3SG
 'Søren made an airplane.'
- (b) Suluusaqarpoq aquuteqarllunilu
 wing-have-INDIC-3 SG rudder-have-INF-4SG-and
 'It has wings and a rudder' (Sadock 1980:311)

Similar phenomena are found in a variety of other languages including Mohawk (Mithun 1984:869). The fact that that the IN is able to serve as the antecedent for a discourse anaphor is strong evidence, in a syntactic approach, that NI is derived syntactically. If it were derived lexically, then the IN would form a single lexical unit with the verb and thus would not be available to serve as a discourse referent (the word being an anaphoric island).

Baker 1988, 1996, Baker et al. 2005 take some kinds of incorporation to be morphological or purely lexical and some kinds to be syntactic. Of Mithun's four types of incorporation (above), Baker et al. 2005 consider types III and IV, and possibly II, as syntactic. Appendix B shows the classifications of Mithun, Baker, and Rosen, and also our classifications in the same terms made when possible. It is not always possible to tell from grammars whether incorporation is syntactic or not, since examples showing the IN as discourse anaphoric referent, with external modifiers, etc. are not always given even where the language can form them, and their non-occurrence is even less often noted. All the languages left unclassified in Appendix B show at least some diagnostic effects that are compatible with a syntactic origin of their incorporation and with a syntactic object function for the IN. These include restriction of incorporation to particular argument roles including object, transparent compositional semantics,⁸ and availability of a non-incorporating paraphrase.

Without taking a stand on the overall nature of NI, we assume it to be uncontroversial that there are some aspects of NI that are the result of derivation in the syntax—defined atheoretically as that component of the grammar that is responsible for deriving relations between words in a sentence. The question then is whether the observed tendency for the ordering of verb and incorporated noun can be attributed to the principles of the syntax.

⁸ Transparent semantics seems to us to be a necessary condition for syntactic incorporation, though it is not a sufficient condition as it could also be compatible with a derivational origin.

2.4. Productive incorporation

We consider a pattern of incorporation to be productive if it can involve any member of a non-closed set of nouns. There are languages that can incorporate only and all body part terms, e.g. Totonaco (Totonac-Tepehua: Mexico) shown in (19), and languages that can incorporate all body part terms and some or many others, e.g. Halkomelem (Salishan: Canada), shown in (20-21). Since body part terms are a large and probably open class of nouns, we count these patterns as productive.

(19) k-laqa-cakaa

1:S-face-wash

'I wash my face'

(T. McFarland fieldnotes)

(20) lək^w-xén

get.broken-leg

'break a leg'

(Suttles 2004:307)

(21) t^θq^w-élwəs-t

punch-side-TRANS

'punch him on the side'

(Suttles 2004:307)

Whether the verbs that allow incorporation are a closed set or not is immaterial to our definition of incorporation. Thus if a language has a closed set of incorporating verbs but an open set of incorporating nouns we consider the incorporation to be productive. An

example is given in (22) from Tümpisa Shoshone (Uto-Aztecan: U.S.), in which the possible incorporating verb stems are limited to just five but these allow any noun to incorporate into them.

- (22) nümmü so'oppüh putisih pungkupaimmippühantü
 we.EX many burro pet-have-HAB-PST
 'We used to have many burro pets' (Dayley 1989:91)

Another language with a closed set of verbs taking incorporation is Warembori (Lower Mamberano: Papua New Guinea), where in one type of incorporation "certain verbs that require the presence of a noun in order to be accomplished may incorporate that noun" (Donohue 1999:45). Also in Warembori, the subject of the existential verbs 'exist' and 'not exist' must be incorporated (*ibid.*).

The verb involved in productive NI must be a real verb. Note that, while for Tümpisa Shoshone the verbs involved in NI are members of a closed class, they are not simply verbalizing suffixes. They can be used independently and combine with the incorporated noun in a semantically transparent manner.

We have defined productivity solely in terms of the nonclosed class status of the incorporated noun, excluding the status of the verb, because of the fundamental asymmetry in the relationship of the constituents of a clause: the verb is the head and so can determine the properties of (subcategorize for) for its complement noun. Thus, it is not unexpected that some verbs might be able to select for a nominal object that will incorporate into them, just as some others might select for a generic object, or an inanimate object.

All cases of syntactic incorporation identified by Mithun or Baker et al. are productive, as are all of the cases that we have identified as types II-IV, with the single exception of Catalan, a puzzling case which belongs to Type II in view of its case-changing impact on the clause (see Appendix A) but is not productive by our definition (though it has enjoyed some diachronic productivity as a compounding type: see Klingebiel 1989:202-224).

3. Methods and results

We searched for languages with noun incorporation, using as leads secondary sources, our own knowledge of languages, the Autotyp database (Bickel and Nichols 2002), and additional surveying of grammars. Where sister languages have incorporation we chose one language per stock or (for older stocks) primary branch (this level is comparable to or slightly older than the *genus* of Dryer 1989, 2005b).⁹ We found 45 languages (representing different families or major subfamilies) with some type of incorporation pattern. Four of the languages have two different kinds of incorporation (one productive or syntactic and one not), for a total of 49 different incorporation patterns, 39 of them productive. The figures and tables below usually count patterns, not languages.¹⁰

⁹ There are four cases in our sample of languages from different branches of stocks: Kwak'wala and Nuuchahnulth, both Wakashan; Bininj Gun-Wok and Ngandi, both Gunwinyguan; Samoan and Tukang Besi, both Austronesian; Sora and Car Nicobarese, both Austroasiatic; and Blackfoot and Cree, both Algonquian.

¹⁰ We are aware of four additional families that have or may have NI, all in South America (page numbers from Adelaar and Muysken 2004): Cahuapanan (449), Chocoan (59), Chonan (563), and

Map 1 shows the distribution of the languages with object incorporation among the 200 languages of the Autotyp genealogical sample that have been surveyed for incorporation. As the map shows, incorporation (and especially productive incorporation) is overwhelmingly a Circum-Pacific phenomenon (as that area is defined by Bickel and Nichols 2006: it comprises the Americas, Oceania-New Guinea-Australia, and eastern Asia back to the major coast mountain range). The only instances outside the area are Catalan, Frisian, and Somali.

3.1. The incorporated noun is generally preverbal

Regardless of type, incorporation exhibits a general tendency for preverbal position of the incorporated noun:

| | | | |
|------|----------|----------|------------------|
| (23) | IN + V | V + IN | Total |
| | 36 (73%) | 13 (27%) | 49 ¹¹ |

Esmaraldeño (159; extinct, unclassified language). Dixon & Aikhenvald 1999 also mention some additional languages from Amazonia, for all of which the IN is preverbal but we cannot determine any other properties of NI. For none of these languages do we have enough information to include them in the survey. As will become clear below, information on productivity of incorporation for any of these would be extremely valuable.

¹¹ If Murrinh-patha and Marrithiyel are classified as V+IN, 35 (70%) and 15 (30%). If Saweru, Car Nicobarese, and Samoan are excluded: IN+V 37 (79%), V+IN 10 (21%). With both of these adjustments, 35 (74%) and 12 (26%).

3.2. The incorporated noun in unproductive NI is nearly always preverbal

In our sample productive incorporation is postverbal in just over one-third of the cases but unproductive incorporation never is, as shown in (24). The difference between productive and unproductive incorporation is categorical in our sample, but only barely statistically significant because of the small sample size, chiefly the small number of languages with unproductive incorporation.

| (24) | IN+V | V+IN | Total | |
|--------------|------|------|-------|---|
| Productive | 26 | 13 | 39 | |
| Unproductive | 8 | 0 | 8 | |
| Unknown | 2 | 0 | 2 | |
| Total | 36 | 13 | 49 | $p < 0.05 (X^2), < 0.06 (\text{Fisher})^{12}$ |

3.3. Position of the incorporated noun correlates with word order in productive NI

For productive incorporation, the position of the incorporated noun is sensitive to the language's word order, as shown in (25). The incorporated noun tends to be on the same

¹² $p < 0.04$ on both tests if the probably unproductive Damana and Retuarã are added.

Significance is improved slightly if Marrithiyel and Murrinh-patha are reclassified as V+N (see Appendix A); it is reduced slightly if Saweru, Car Nicobarese, and Samoan are excluded. The validity of the chi square test is dubious because of the small sample size (the zero cell contributes disproportionately to the significance), so significance levels reported below are based only on Fisher's Exact Test unless otherwise stated.

side of the verb as an unincorporated object and the difference in frequency of postposed incorporated nouns is highly significant. A stronger tendency, though, is the preference for preverbal position, which overrides basic word order in about half of the VO languages.

(25) Productive incorporation and word order. n.d. = no data (on word order).

| | IN+V | V+IN | Total | |
|-------|----------|----------|-------|--|
| OV | 15 | 2 | 17 | |
| VO | 9 | 10 | 19 | |
| Free | 1 | 0 | 1 | |
| n.d. | 2 | 0 | 2 | |
| Total | 27 (69%) | 12 (31%) | 39 | p < 0.02 (OV/VO), p < 0.03 (OV/other) ¹³ |

If syntactic derivational history were responsible for the ordering of the incorporated noun and verb, productive incorporation should be no less, and probably more, consistently preverbal than unproductive incorporation, and should be less influenced by the language's basic word order. Though the number of languages in our sample with unproductive NI is too few to reach statistical significance or allow firm conclusions, it appears that, on the contrary, unproductive incorporation is more often preverbal, as shown in (26). The difference in the frequency of V+IN incorporation in VO languages is telling: V+IN is in a

¹³ X² (valid) p = 0.024. Same or better significance if Samoan, Saweru, and Car Nicobarese et al. are excluded and/or if Murrinh-Patha is reclassified as V+N (calculated for OV/other only).

slight majority for productive incorporation (25) but absent for unproductive incorporation (26).¹⁴

(26) Unproductive incorporation and word order

| | IN+V | V+IN | Total | |
|-------|------|------|-------|-----------------------------|
| OV | 3 | 0 | 3 | |
| VO | 4 | 0 | 4 | |
| Free | 1 | 0 | 1 | |
| Total | 8 | 0 | 8 | n.s. (p=1.00) ¹⁵ |

3.4. Dryer's test

The continent-by-continent test of Dryer 1989 is a simple and reliable way of way of testing for significance of a distribution or correlation based on whether it holds in every continent, and how strongly. Unfortunately, the incidence of noun incorporation is so low overall and especially in Eurasia and Africa that getting appreciable representation of all six of Dryer's continent-like areas, and meaningful margins within each, is nearly impossible. Still, by lumping low-incidence continents together one can see whether preferences obtain worldwide. Appendix E shows several patterns and correlations across four large areas: Africa and Eurasia; the Pacific (Australia, New Guinea, Oceania), North

¹⁴ The difference for just these four datapoints (IN+V/V+IN for VO in (25) and (26)) is nearly significant (p=0.081) despite the tiny sample size.

¹⁵ Also non-significant if Marrithiyel unproductive NI is reclassified as V+N.

America (including Mexico and all of the Mesoamerican cultural and linguistic area), South America (including Panama). Productive NI is more frequent than unproductive in all four areas. Preverbal IN is more common than postverbal in all four areas. The correlation of productivity and IN order obtains in all four areas, unsurprisingly as the incidence of unproductive NI and V+N incorporation is zero everywhere as shown above. The correlation with word order obtains in that three out of four areas have good representation of both N + V and V + N order for VO languages and all four have zero or near-zero frequency of OV with V+N order. All of this is further evidence that the preferences we have identified are universal.

4. Discussion

Two previous works have found a correlation between IN order and word order: Mardirussian 1975 and Kozinsky 1981. Both find two-way implicational correlations with the order of subject and verb (Universals 358-359 and 1492-1493 in Plank et al. n.d.):

(27) Mardirussian: Verb-initial implies V+N, non-verb-initial implies N+V

Kozinsky: N+V implies S(...)V, V+N implies V(...)S¹⁶

¹⁶ Kozinsky 1980 is cited from Plank et al. n.d.; the original was not available to us.

Both used smaller and/or less consistently genealogical samples than ours (i.e. had fewer genealogically independent datapoints). Mardirussian defines incorporation more broadly than we do so that it includes some of what we consider N+V compound verbs (§4.2) and/or DOM. Perhaps these differences explain why both sources find two-way correlations while we ours can be stated as a one-way correlation:

(28) V+N order in incorporation implies VO clause order

(The result is similar if we use the order of subject and verb, as Mardirussian and Kozinsky did. It is clearer for object and verb, if only because we have a few more data gaps for SV/VS order than for OV/VO order.)

It is not a categorical universal that the incorporated noun always precedes the verb; (23) above shows that over one-quarter of our sample languages have V+N order. As a noncategorical generalization, however, it is quite strong, and we propose that it should be recognized as a typological universal:

(29) *Baker's universal*

In noun incorporation, the incorporated noun tends to precede the verb.

The fact that it is most evident in unproductive incorporation ((26)) cannot be explained by syntactic derivation, whose effects should be clearest in productive incorporation. Where the syntax establishes a relationship between the verb and the incorporated noun, it can be expected that the linear order of these two elements is determined by the same processes or

rules that give linear order to other syntactic constituents, in particular the verb phrase. The position of the incorporated noun should then always parallel the position of unincorporated object noun phrases. Furthermore, incorporation might result diachronically from univerbation of an adjacent noun and verb, and such cases would have the language's regular word order. As shown in §3 above, however, ordering in incorporation and in clause syntax are not always identical; there is a clear preference for preverbal incorporated nouns, regardless of word order, especially in unproductive incorporation. That is, in fossilized or more nearly fossilized contexts, ordering is not determined by the syntax and must be determined by something else. We believe it is a constraint or preference, perhaps morphological, whereby languages tend to place the noun before the verb, or the nonhead before the head, in certain kinds of word structures. To test this we have done cross-linguistic surveys of three other kinds of noun-verb combinations that are common enough to be easily surveyable.

4.1. Synthetic compound nouns

Synthetic compound nouns (for this term and its history see Bauer 2001) are compound nouns containing a noun root and a verb root, where the noun is the semantic object of the verb. Examples are English *skyscraper*, *witch hunt*, and *scarecrow* or Spanish *matamoscas* 'fly swatter' (lit. 'kill-flies') and *lavaplatos* 'dishwasher' ('wash-plates'). There seem to be three ways in which such compounds can be constructed: noun-noun compound consisting of noun plus nominalized verb; nominalization of an entire

noun+verb compound stem; and zero-derivation or conversion of a noun + verb sequence.

The three types are illustrated in (30)-(32) (NZ = nominalizing morphology).

- (30) N + V-nz Ingush (Nakh-Daghestanian, Caucasus; JN fieldnotes)

chq'earii+du'-arjg

fish-PL + eat-NZ 'heron, stork'

(likewise English *skyscraper*, etc.)

- (31) [N + V]-nz Nisgha (Tsimshianic, British Columbia; Tarpent 1999:795)

ha - [yò'oks + 'wé:n – tkw]

INSTR wash.s. + teeth-MEDIAL 'toothbrush'

- (32) [N + V]_N Abkhaz (West Caucasian, Georgia; Chirikba 2003:27)

a - c'la + r-k'wók'w

ART-tree + [CAUS-split] 'woodpecker'

(likewise English *scarecrow*, etc.)

Assignment to one or another subtype is not particularly important for our purposes; the point is that all three illustrate the single phenomenon of synthetic compound noun containing a noun root and a verb root. We sought examples of such compounds, surveying a genealogically and geographically distributed sample of 52 languages (some of them taken from the similar surveys of H. Anderson 1997 and Bauer 2001). Anderson found, and we also found, that OV languages have almost exclusively N+V order in such

compounds, while VO languages have either V+N (following the basic clause word order) or N+V. Bauer found that the ordering in these compounds responds more to morphological than to syntactic tendencies, in many languages following the order in other compounds rather than the clause word order, and in fact his morphologically-driven departures from syntactic word order are all cases of VO languages with N+V compounds. Our findings are given in Appendix C and summarized in (33). There is a highly significant correlation between word order and compound-internal order, but there is also a very strong preference for preverbal position regardless of word order.

(33) Ordering in synthetic compound nouns¹⁷

| | N + V | V + N | TOTAL | |
|-----------|-------|-------|-------|--|
| OV | 26 | 0 | 26 | |
| VO | 8 | 15 | 23 | |
| free/none | 3 | 0 | 3 | |
| TOTAL | 37 | 15 | 52 | p < 0.0000003 (OV vs. VO); p < 0.0000003 (OV vs. other) |

¹⁷ While native Japanese compounds have N+V order, corresponding to OV word order, e.g. *hito-gorosi* 'person-killing; manslaughter', Sino-Japanese compounds use V+N order, e.g. *satuzin* 'kill-person; manslaughter' (Shibatani 1990:240-241). Since both elements of these compounds, and often the compound itself, come from Mandarin Chinese, we do not include this V+N pattern in our counts.

4.2. Nonsyntactic incorporation and compound verbs.

The same situation occurs with clearly nonsyntactic object incorporation (Mithun's Type I) and similar lexical compounds (these differ from the synthetic compound nouns only in that they are verbs). The sample languages are shown in Appendix B, plus Choctaw and Itelmen from Mithun 1984 and Kutenai (isolate; Matthew Dryer, p.c.) from our additional survey.

(34) Ordering in nonsyntactic incorporation and compound verbs.

| | N + V | V + N | TOTAL |
|-----------|-------|-------|-------|
| OV | 9 | 0 | 9 |
| VO | 3 | 3 | 6 |
| free/none | 1 | 0 | 1 |
| TOTAL | 13 | 3 | 16 |

$p < 0.044$ (OV vs. VO;

$p < 0.036$ (VO vs. other)

4.3. Instrumental affixes.

Instrumental affixes are derivational elements that indicate the instrument or means used to accomplish the action of the verb. Most commonly they indicate "action with the hands ... with the feet ... with the mouth ... with the buttocks or body weight ... natural forces ... and actions with certain kinds of instruments, especially long ones, compact ones, and sharp ones..." (Mithun 1999:126). They are found in many language families and

isolates of North America, including Siouan-Catawba, Uto-Aztecan (Numic branch), Yuman, Chumash, Pomoan, Palaihnihan, Chimariko, Shasta, Maidu, Wappo, Washo, Klamath, Takelma, Sahaptian, Kutenai, Haida, and Algonquian (cited by Mithun 1999:118-126). They are also found in Kutenai (isolate; Canada and U.S.), the Lule-Vilela family of Argentina (Adelaar and Muysken 2004:387), Chayahuita (Cahuapanan, Peru; Adelaar & Muysken 2004:448), and Oceanic (Austronesian) languages of New Caledonia (e.g. Tinrin: Osumi 1995:120-123). We surveyed all these languages, which were mentioned in secondary sources, and also sought instrumental affixes in a larger worldwide sample (which yielded no more examples). We looked for (1) a more or less dedicated affix or compound-element slot with (2) a number of fillers having body part meanings and (3) primarily instrumental meaning for those elements.¹⁸ In every language found, with the exception of the Algonquian family and Kutenai, instrumental affixes are prefixes. Figures are shown in (35); see Appendix D.

Although "the origins of most instrumental affixes are no longer recoverable" (Mithun 1999:123), for some families reconstruction has traced them to noun roots (M. Nichols 1974 for Numic, Leer 1977:95 for Haida), and the fact that many of them indicate

¹⁸ A few languages have an occasional body-part meaning in an affix slot containing other elements, or an occasional instrumental meaning in a body-part element. E.g. Ngiyambaa (Pama-Nyungan, Australia: Donaldson 1980:202-3) has a set of bound verbs used in compounds, two of which are glossed with body part terms but neither of which is instrumental in actual functions; Tawala (Austronesian: Oceanic, New Guinea; Ezard 1978) has a compound slot dedicated to body part terms which are not instrumental (222-226) and a prefix slot with five elements one of which is glossed 'hand' or 'persistent action' (175, 178-9) and is instrumental when glossed 'hand'. Neither of these languages meets our three survey conditions.

body parts makes them inherently noun-like in their semantics. To the extent that they originate from nouns or pattern synchronically with nouns, their position relative to the verb illustrates our generalization.

(35). Instrumental affix position and word order. Count is by families where all daughters surveyed have the same word order; where they differ, each pattern is counted separately.

(See Appendix D.)

| | Ins + V | V + Ins |
|-------|---------|---------|
| OV | 10 | 0 |
| VO | 6 | 1 |
| Free | 3 | 1 |
| Total | 19 | 2 |

5. Conclusions

We have shown that incorporated nouns generally precede the verb root regardless of the language's clause word order, that this tendency is strongest in unproductive noun incorporation, and that it is even stronger in synthetic compounds and instrumental affixes. That is, productive incorporation is somewhat likely to be structured by language's syntax (and indeed usually has non-incorporation, i.e. free syntactic construction, as a readily

available paraphrase), while unproductive incorporation is more likely to follow the cross-linguistic preference we have found, and purely lexical compounding even more so.

Small sample size has made it difficult to reach high statistical significance in our counts. It is unlikely that the sample can be expanded appreciably, as our survey has probably found all known cases of noun incorporation at least at the genus level and has furthermore defined incorporation rather generously. There are two ways in which the findings might be made firmer. One is to expand the survey to include more sisters of our sample languages and apply a randomization test like that of Bickel et al. 2006.

Another would be to increase the information available on languages known to have noun incorporation, particularly information on productivity of incorporation and/or position of IN in the additional languages from South America mentioned in footnote 10.

We have described the cross-linguistic preference as positioning noun and verb roots, but it could be that the relevant factor is not part of speech but head vs. nonhead, with heads favored to follow nonheads (Williams 1981 formulated this preference as the Righthand Head Rule, offered chiefly for English). Possible support for this is the fact that for three cases of verb+noun compounds where we have information on centricity the verb+noun type is exocentric, i.e. the verb is not the head: the Slavic compounds mentioned just below, the English minority type of *scarecrow*, *breakwater*, etc. (Jespersen 1948:262-3, Marchand 1969:15-16, 380-381, Kastovsky 2006:169), and the highly productive Romance type of Spanish *matamoscas* 'flyswatter'. These compounds differ from the noun+verb ones that conform to our universal not in the parts of speech of their components but in their centricity.

Little is known about how the asymmetry between productive and unproductive incorporation arises. We know of no cases where incorporated nouns have moved from postverbal to preverbal position. There is a telling case in Halkomelem (Salishan, British Columbia), where incorporation is almost entirely postverbal but there are a few preverbal incorporates (recall that the Salishanist terminology is *lexical suffix*, *lexical prefix*). One preverbal incorporate may have originated when an indelicate postverbal incorporated noun was deleted, whereupon a local prefix was reinterpreted as the incorporated noun (Galloway 1993:200-201 for the Upriver variety, Suttles 2004:282 for the downriver Musqueam variety). That is, no incorporated noun moved; rather, when the semantics demanded an incorporated noun but the form did not provide one, a prefix was reanalyzed as the incorporate, going against the syntactic order of the language and producing a preverbal incorporate.

Also revealing is the case of Frisian, which has developed productive non-syntactic incorporation by back-formation from synthetic compounds (analogous to the English verb *cherry-pick* from *cherry-picking*). Dijk 1997 argues that this was made possible by the appropriate nonfinite morphology (analogous to English *-ing*) and OV word order, and indicates that incorporation arises naturally in Germanic varieties that meet these two conditions. Here it is not that compounding is frozen former syntax, but that preverbal incorporation arises by reanalysis of compounding.

Another telling case involving compounds comes from Russian and other Slavic languages (Kiparsky 1975:345-350, Progovac 2008). These languages have unproductive verb + noun exocentric compounds (Russian *sorvi+golova* [tear off + head] 'madcap, daredevil', *Vladi+vostok* [rule + east], Bosnian/Croatian/Serbian *kljuj+drvo* [peck+wood]

'woodpecker') which descend from Indo-European antecedents, as well as productive noun + verb endocentric compounds (e.g. Russian *jazyk-o+znanie* [language-LINK+knowledge] 'linguistics', *posud-o+moj-ka* [dishes-LINK+wash-SUFFIX] 'dishwasher') ultimately based on models calqued from Greek in the middle ages. No nominal element moved to produce the second type, but rather a borrowed model was taken up and has become productive though its ordering goes against the syntactic word order, which is VO.

A fourth case comes from Ngan'gitjemerri (Western Daly: Australia; a very close sister to our survey language Murrinh-patha), for which Reid 2003 presents evidence of diachronic change. Early 20th-century Ngan'gitjemerri used rather freely ordered combinations of light verb and heavy non-verbal piece, and by the late 20th century these had solidified into complex verb stems with a preposed conjugating auxiliary and a postposed lexical root. Object noun incorporation is absent in early texts. As the former light verb was reanalyzed as the conjugation and the language became polysynthetic, the light verb came to be preposed to the heavy piece and the former object of the light verb was reanalyzed as prefixed to the heavy piece, now the lexical verb root (Reid 2003:110-112). Again, there was no movement of the IN to achieve the desired order, but rather reanalysis of a noun as incorporated once it occurred in the right position.

Finally, in Catalan, the N+V order of incorporation does not derive directly from early Latin or Indo-European word order and is not the result of univerbation, but arose in the written language as an extension of existing word-formation patterns in the late middle ages (Klingebiel 1989).

It has long been natural to explain the internal ordering in compounds and complex verb forms as unverbated syntax freezing the word order at the time of univertation;¹⁹ specifically concerning compounds, Lehmann (1969; later summary 1975) argued that Germanic N+V synthetic compounds are an archaism reflecting OV order in early Indo-European. In fact, though, in none of the cases reviewed above, the only ones for which we are aware of diachronic evidence, is the ordering in incorporation due to simple frozen syntax. In Halkomelem and Slavic conformities result from selection of morphological models; in Frisian and Catalan there was no univertation but adaptation and expansion of existing morphological patterns; in Ngan'gitjemmerri there was univertation but the ordering of elements in incorporation results from reanalysis and does not straightforwardly reflect the earlier clause word order.

Thus, unless all examples known to us happen to be atypical, the ordering of elements in unproductive incorporation is not a simple reflex of univertation. Nor could univertation per se account for a strong cross-linguistic preference for one order over another (unless there is a strong preference to change from OV to VO but not the reverse, so that the lack of OV languages with V+N compounding is due not to principles of compounding but to an absence of OV languages that were formerly VO). Similarly, synchronic consistency between clause word order and morpheme order in verbs is known to obtain cross-linguistically (e.g. Siewierska & Bakker 1996) but cannot account for the asymmetry we have found.

If the preference for N + V incorporation and compounding does not reflect frozen word order or synchronic word order, neither does it appear to reflect any preference for

¹⁹ See Siewierska & Bakker 1996:note 23 for the long history of this claim about univertation.

particular diachronic processes, as the kinds of reanalyses that occurred in Halkomelem and Ngan'gitjemerri are different, and the use of existing morphological models in Frisian and Catalan was different. Rather, the traceable diachronic cases all seem to be driven by the outcome itself: they have evolved as though they were the result of selection of models in favor of verb-final or head-final order, in most cases overriding the syntactic word order.

Is it plausible that verb-final or head-final order might be a universal background or default preference? It is known that verb-final word order is preferred cross-linguistically (Dryer 1989), a preference still in need of explanation.¹ As an alternative explanation, argument agreement markers are more likely to be preposed than other inflectional markers on verbs (Bybee et al. 1990, Enrique-Arias 2002), so perhaps IN's -- which are arguments in some or all respects -- are following the same pattern as affixes. Arguing against this account are the fact that incorporation is not affixation and the fact that affixal agreement marking is not absolutely likely to be prefixal but only more likely than other inflectional categories are, while IN's are absolutely likely to be preverbal. In favor of a universal head-final background default is the conclusion of Pycha 2008 that, in phonology, the rightmost element in a morphologically complex word tends to be the strongest, controlling such processes as reduplication, umlaut, vowel elision, etc., and that the head-final preference in incorporation and compounding is an instance of the same very general strength relationship. Since universal head-final ordering has some support, is interestingly contentious, and is easy to invoke but hard to demonstrate, we propose it as a hypothesis for future work:

¹ Newmeyer 2000 suggests that SOV order may be an archaism inherited from Proto-World, but Maslova 2000 shows that this is unknowable on mathematical grounds.

(36) Hypothesis: Head-final ordering is the default.

It can be falsified by showing, language by language or pattern by pattern, that ordering can be better accounted for by other factors. It can be supported, language by language, by falsifying its negation (e.g. showing that it can't be that the language has otherwise inexplicable head-final cases remains) or by showing that an otherwise inexplicable residue of head-final cases remains after an otherwise complete analysis. In this paper we have shown that the facts of incorporation make the hypothesis plausible. That IN's are usually preverbal is also a strong and useful typological generalization even if the hypothesis fails more generally.

University of California, Berkeley

Appendix A: Examples from languages with NI

Examples of NI from languages in our sample that were not included in the text follow. The incorporated noun in each example is boldfaced; discussion is included only as is necessary.

Ainu (isolate: Japan and Sakhalin Islands)

- (37) kane rakko o-**tumi**-osma
 golden otter APP-war-begin
 ‘The war started because of the golden sea otter.’ (Shibatani 1990:63)

Alamblak (Sepik Hill: New Guinea)

- (38) wa-**yufa**-yuta-n-r
 IMPER-name-call-2SG-3SGM
 ‘Call (his) name!’ (Bruce 1984:170)

Bininj Gun-Wok (a.k.a. Mayali) (Gunwinyguan: Australia)

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

Caddo (Caddoan: southern U.S. plains)

- (40) hak#ku-**nas**-sininih-sa’
 IND#1P-foot-tingle-IMPF
 ‘My foot is asleep.’ (Melnar 2004:44)

Catalan (Indo-European, Romance: Spain)

- (41) El caçador va **cama**-trenc-ar l’ocell
 the hunter PST leg-break-3SG.PST the.bird
 ‘The hunter broke the bird’s leg(s).’ (Gràcia and Fullana 1999:240)

Chukchi (Chukotko-Kamchatkan: Chukotka, Russia)

- (42) taŋ-amənan Cəkwaŋaqaŋ ya-**qora**-nm-at-len
 INTS-alone personal.name.3SG.ABS PERF-reindeer-kill-VB-3SG
 ‘Cəkwaŋaqaŋ all by himself slaughtered reindeer.’ (Dunn 1999:222)

Cree (Algic: Canada)

- (43) kaskih**ci**kwaṅēhwēw
 ‘He breaks his knee by shot.’ (-ihci**kwān**- ‘knee’) (Wolfart 1973:67)

Damana (Chibchan: Colombia)

- (44) **suzu-n-Ø-go-ka**
backpack-CONNECTIVE-3SG-knit-FACTUAL
'She knits backpacks.' (Quesada 1999:248)

Frisian (Indo-European, Germanic: Netherlands)

- (45) Wat be-**popke**-teken-est de hiele tiid?
what PREFIX-figurine-draw-2SG the whole time
'Why the hell are you drawing figurines all the time?' (Dijk 1997:19)

Guaraní (Tupi-Guaraní: Paraguay)

- (46) (Che) ai-**po**-pete la-mita
I 1:ACC-hand-slap DEF-child
'I slapped the child in the hand.' (Velazquez-Castillo 1996:99)

Haida (isolate: western Canada)

The formative *ts'a* is part of the verb stem, preceding the incorporated noun *da sq'asgiid*.

- (47) giyaangw-ee 'la ts'a **da sq'asgiid** sdang-gan
cloth-DEF 3PL INSTR yard be.two-PAST
'She cut a two-yard length of the cloth.' (Enrico 2003:787-90)

Kiowa (Kiowa-Tanoan: U.S. Great Plains)

- (48) é-**ál̀:̣p-è:**
2/3SGA:1SGP:SGO-apple-request:IMPF
'He asked me for an apple.' (Watkins 1984:226-227)

Kwak'wala (Northern Wakashan: western Canada)

- (49) q'əmdzək^w-ila-iḡsd-ida bəg^wanəma-ḡa q'əmdzək^wi?
salmonberries-give.feast-want-DET man-OBJ salmonberries
'[when] the man wants to give a salmonberry feast (of salmonberries)'
(Anderson 1992:30)

Lakhota (Siouan-Catawba: U.S.)

An incorporated body part noun cannot take possessive prefixes, even if obligatorily (inalienably) possessed as an independent noun:

- (50) **napé** makipozo we
hand 1:S-DAT-show IMP
'Show me your hands.'

- (51) *ninapé makìpozo we
 2:S-hand 1:S-DAT-show IMP
 ‘Show me your hands.’ (de Reuse 1994:210)

The following show a locative prefix optionally preposed to the entire NI compound:

- (52) **châ**-tagle
 heart-set
 ‘have designs on someone’
- (53) a-**châ**-tagle
 LOC-heart-set
 ‘have designs on someone’ (de Reuse 1994:213)

Marrithiyel (Western Daly: Australia)

Nonsyntactic, unproductive NI:

- (54) Ngirringgi-**yan**-dim-Ø-a
 1PLEXS.RR.IRR-nose-sink-PLS-PAST
 ‘We should have drowned him.’ (Reid 2003:117-19, following Green 1989)

Syntactic, productive NI:

- (55) Ginj-inj-duk-**miri**-ya sjiri
 3SGS.NJ.PF-2SGO-pull.out-eye-P splinter
 ‘She removed a splinter from your eye.’ (Reid 2003:117-19, apud Green 1989)

These are the only examples cited, and ‘eye’ in the productive example is not clearly object, but Reid (2003:117) says that this type does include object incorporation (RR, NJ = roots of conjugating auxiliaries). Reid’s terms for the two types are *lexical* and *syntactic* respectively. In the unproductive type the IN precedes the lexical or heavy verb piece (Reid: coverb) and follows the conjugating auxiliary or light verb (Reid: finite verb). We take the conjugating auxiliary to be just a bearer of inflectional marking, so that unproductive incorporation is preverbal. Footnotes in section 3 track the implications of changing this analysis and taking the auxiliary to be the verb.

Movima (unclassified, probably isolate: Brazil)

Examples showing classificatory bound root incorporated into the verb.

- (56) loy iʔ chuʔ-na as manka
 ITN 1 knock.down-DF ART.N mango
 ‘I’ll pick the/a mango.’ (Haude 2006:283-4)
- (57) loy in’ chuk-a:-**ba** n-is manka
 ITN 1INTR knock_down-DR-BR.round OBL-ART.PL mango
 ‘I’ll pick mangos.’ (Haude 2006:283-4)

- (58) *loy iɬ chuk-a:-**ba** is manka
 ITN 1 knock.down-DR-BR.round ART.PL mango
 (Haude 2006:283-4)

Incorporated full noun with external coreferential noun:

- (59) n-os chon' pul-a-**lolos**-wa-y'hi n-os lo:los
 obl-ART.N.P HAB sweep-DR-yard-NMZ=1P OBL-ART.N.P yard
 '...when we always swept the yard.' (Haude 2006:284)

Murrinh-Patha (Southern Daly: Australia)

- (60) thunku dem-ngi-**darri**-lerrkperrk
 fire 3A-1O-back-heat
 'The fire makes my back feel hot.' (Walsh 1987; retranscription)

This NI construction admits the same two analyses as the Marrithiyel one above, and is similarly tracked in footnotes.

Nambikwara (isolate: Brazil)

- (61) Pre²ka²ki³a² ũ³-**ha³la³**-kwã¹t-a¹-hẽ³-la².
 nail INSTR-NI:middle-bend-1SG-T/E:IO:PAST-PERF
 'I bent the nail near the middle.' (Kroeker 2001:67)

Ngandi (Gunwinguan: Australia)

- (62) ngagu-**jundu**-geyk-dh-i gu-jundu-yung
 1SG>3SG-stone-throw-?-TNS NCM-stone-ABS
 'I threw a stone.' (Heath 1978:118-19; retranscription)

Nisgha (Tsimshianic: British Columbia)

- (63) qùɬ-**hó:n** n`ì:y
 gut.something-fish me
 'I gutted fish.' (Tarpent 1987:792)

Nuuchahnulth (Wakashan: British Columbia)

(64) is an unincorporated clause showing regular verb-initial word order. The verb is a bound verb which cannot occur independently; to be used without incorporation it must take the referential stem (REF), a dummy incorporated element. (65) has incorporation. Though at first glance the incorporated noun *suuḥaa* 'spring silver salmon' does not appear to be flanked by verbal material, the absence of the referential stem and the otherwise impossible OV word order make it clear that the first element in (65) is incorporated.

- (64) ?u-ityaap-'atl suuḥaa
 REF-bring.as.gift spring.silver.salmon
 'He brought a gift of spring silver salmon.'

- (65) **suuḥaa**-ity'aap-'at
spring.silver.salmon-bring.as.gift-PASS
'He brought a gift of spring silver salmon.'
(Stonham 2004:219)

Oneida (Iroquoian: northeastern U.S.)

- (66) la-**ahy**-k-s sá.yes
PRO-fruit-eat-SER blackberries
'He's (fruit-)eating blackberries.'
(Abbott 2000:63)

Retuarã (Tucanoan; Colombia)

In (67) ditransitive 'put' has become monotransitive as a consequence of incorporation (*sa-* agrees with 'canoe' and not with incorporated *hãã* 'seat').

- (67) bikitoḥo sa-ii-teri-**hãã**-râyũ
morning 3.N.SG-3.M.SG-seat-put-FUT
'In the morning he will put seats in it (canoe).' (lit. 'he will seat-put it')
(Strom 1992:100)

Saweru (West Papuan: New Guinea)

- (68) ruama wo-mo mo=**rama** a-bai
woman 3SGF.ERG 3SGF.NOM=man 3SGM-hit
'The woman hit the man.'
(Donohue 2001:327)

Slave (Athabaskan: Canada)

- (69) te-**fí**-yé-njshu
water-head-CONJ-2SG-put
'You have put your head in water.'
(Rice 1989:659; interlinear added)

Sora (Austroasiatic: Munda, India)

- (70) gđ-**bje**-te-ji
cut-buffalo-TNS-3PLS
'They are slaughtering the buffalo.'
(Zide 1997)

Takelma (isolate: Oregon)

- (71) gwen-**waya**-sgō^{ut}-hi
neck-knife-cut-he.them
'He cut their necks off with his knife.'
(Sapir 1922)

Tiwi (isolate: Australia)

- (72) ji-məni-**kəi**-ŋa
he-me-hand-grab
'He grabbed me by the hand.'
(Osborne 1974:47)

Tukang Besi (Austronesian, Malayo-Polynesian: Indonesia)

- (73) No-sai **kuikui**-mo
 3:R-make cakes-PERF
 ‘S/he has made cakes.’ (Donohue 1999a:168)

Warembori (Lower Mamberamo, Irian Jaya: New Guinea)

Syntactic, productive NI:

- (74) e-mune-**mena**-ro
 1SG-kill-dog-IND
 ‘I killed a dog.’ (Donohue 1999b:43-47)

Nonsyntactic, unproductive NI:

- (75) e-**pue**-kambi
 1SG-pig-hunt
 ‘I hunt for pigs’, ‘I go pig-hunting.’

Wari’ (Chapakuran: Brazil)

- (76) Hu **capam’** in rain pacun!
 blow cornbread completely 2SG.RF-3N stone
 ‘Turn the stones into cornbread!’ (Everett and Kern 1997:386)

Washo (isolate: Nevada)

One of the bipartite stem types consists of a body part lexical prefix or (if there is no prefix in the needed sense) an incorporated noun plus one of a closed set of verb stems. The body part is generally intransitive subject or direct object of the verb. In Jacobsen’s morphophonemic transcription the two morphemes are separated by a space.

- (77) dul[~] išl /duléšil/
 hand give
 ‘offer one’s hand to someone’ (Jacobsen 1980:93)

Yana (isolate: California)

- (78) mic’-**au**-wilmi-si-ndja
 hold-fire-on.one.side-PRES-1SG
 ‘I hold fire in one hand.’ (Sapir 1911:269)

Yimas (Lower Sepik: Papua New Guinea)

- (79) ura-mpu-na-**akpi**-api-n
 fire:O-3PL:A-DEF-back:NCM:SG-put.in-PRES
 ‘They are putting (their) backs to the fire.’ (to warm themselves)
 (Foley 1991:320)

Yucatec (Mayan: Guatemala)

- (80) k in ch'ak-**che'**-t-ik in kòol
 TA 1SG:A chop-wood/tree-TRANS-IMPF my cornfield
 'I clear my field.'

(Bricker 1978:16)

Appendix B. Languages with noun incorporation, their types, and examples. *

Classification from Mithun 1984 or Baker et al. 1985. ** Our classification. ***

Author's classification.

| Language | Syntactic | Productive | Mithun type | Rosen type | IN Position | Word order | Example |
|----------------|-----------|------------|----------------|---------------|----------------|---------------|-----------|
| Ainu | 1 | 1 | III* | Class | Pre | OV | App. A |
| Alamblak | 1 | 1 | | Class** | Pre | OV | App. A |
| Bininj Gun-Wok | 1 | 1 | IV | Class | Pre | free | App. A |
| Blackfoot | | 1 | II | | Pre | VO | App. A |
| Car Nicobarese | 1 | 1 | | Class** | Post | VO | (6) |
| Catalan | 1 | 0 | II** | Class** | Pre | VO | App. A |
| Chukchi | 1 | 1 | III | Cpd | Pre | OV | App. A |
| Cree | 1 | 1 | | | In/Post | n/a | App. A |
| Damana | 0 | 0? | | Cpd** | Pre | OV/free | App. A |
| Frisian | 0 | 1 | I, II*** | Cpd*** | Pre | OV | App. A |
| Haida | 0 | 0 | | Cpd ** | Pre | OV | App. A |
| Halkomelem | 0 | 0 | | Cpd*** | Pre | VO | |
| Halkomelem | 1 | 1 | | Cpd*** | Post | VO | (20)-(21) |
| Kiowa | 1 | 1 | | Class** | Pre | OV | App. A |
| Kwak'wala | 1 | 1 | | Class** | Pre | VO | App. A |
| Lakhota | 0 | 1 | | | Pre | OV | App. A |
| Mapudungun | 1 | 1 | III* | Cpd ** | Post | VO | (2) |
| Marrithiyel | | 1 | | | Post | OV | App. A |
| Marrithiyel | 0 | 0 | | | Pre/Post | OV | App. A |
| Movima | 0 | 1 | I*** | Class** | Post | VO | App. A |
| Murrinh-Patha | 1 | 1 | | | Pre/Post | OV | App. A |

| | | | | | | | |
|------------------|---|----|------|-----------|------|------|-----------|
| Nahuatl | 1 | 1 | III | Class** | Pre | VO | (15)-(16) |
| Nambikwara | 0 | 1 | | | Pre | OV | App. A |
| Ngandi | 1 | 1 | IV | Class** | Pre | OV | App. A |
| Nisgha | 0 | 1 | I ** | Cpd ** | Post | VO | (13) |
| Nuuchahnulth | 0 | 0 | | | Pre | VO | App. A |
| Nuuchahnulth | 1 | 1 | | | Pre | VO | App. A |
| Oneida | 1 | 1 | IV | Class | Pre | VO | App. A |
| Retuarã | 0 | 0? | | Cpd** | Pre | OV | App. A |
| Samoan | 1 | 1 | I | Cpd | Post | VO | (7)-(12) |
| Saweru | 1 | 1 | | | Pre | OV ? | App. A |
| Slave | 1 | 1 | | Class?*** | Pre | OV | App. A |
| Somali | 1 | 1 | | | Pre | OV | App. A |
| Sora | 1 | 1 | III | | Post | OV | (1) |
| Takelma | 1 | 1 | | | Pre | OV | App. A |
| Tiwi | 1 | 1 | III | | Pre | VO | App. A |
| Totonac | 1 | 1 | | | Pre | VO | (19) |
| Tukang Besi | 1 | 1 | | Cpd ** | Post | VO | App. A |
| Tümpisa Shoshone | 0 | 1 | I ** | Class** | Pre | OV | (22) |
| Warembori | 1 | 1 | | | Post | VO | App. A |
| Warembori | 0 | 0 | | | Pre | VO | App. A |
| Wari | 1 | 1 | | | Pre | VO | App. A |
| Washo | 0 | 0 | | | Pre | OV | App. A |
| West Greenlandic | 1 | 1 | | Cpd ** | Pre | OV | (18) |
| Yana | 1 | 1 | | | Post | VO | App. A |
| Yeli Dnye | 1 | 1 | | Cpd ** | Pre | OV | (4)-(5) |
| Yimas | 0 | 0 | | | Pre | free | App. A |
| Yucatec | 1 | 1 | II | Class?*** | Post | VO | App. A |

Appendix C. Language data for ordering of elements in synthetic compound nouns (sorted by clause word order, then compound order). AA = Afroasiatic, NC = Niger-Congo. Entries taken from Anderson 1997 have not been checked. Clause order from Dryer 2005 where sources give no basic order. * = The predominant order for this language; the other occurs as a minor or restricted type. Bare initials in Source column = one of the authors.

| Language | Family | Clause word order | Compound order | Source(s) |
|------------------|-------------------|-------------------------|-------------------|----------------------|
| Dongolese Nubian | East Sudanic | OV | N+V | Armbruster 1960:155 |
| Amharic | AA: Semitic | OV | N+V | Leslau 1995 |
| Somali | AA: Cushitic | OV | N+V | Saeed 1999 |
| Kanuri | Saharan | OV | N+V | Anderson 1997 |
| Koyraboro Senni | Songhai | OV | N+V | Heath 1999:170 |
| Sumerian | Sumerian | OV | N+V | Karahashi 2000 |
| Ingush | Nakh-Daghestanian | OV | N+V | JN |
| Abkhaz | West Caucasian | OV | N+V | Chirikba 2003 |
| Georgian | Kartvelian | OV | N+V | various dictionaries |
| Basque | isolate | OV | N+V | Saltarelli 1988:262 |
| Turkish | Turkic | OV | N+V | Anderson 1997 |
| Persian | Indo-European | OV | N+V | MT |

| | | | | |
|------------------|------------------|----|-------|--|
| Paite | Tibeto-Burman | OV | N+V | Singh 2006 |
| Burmese | Tibeto-Burman | OV | N+V | Soe 1999:22-25 |
| Japanese* | isolate | OV | N+V | Shibatani 1990:240-241 |
| Nivkh | isolate | OV | N+V | Mattissen 2003, Gruzdeva 1998:23 |
| Alamblak | Sepik Hill | OV | N+V ? | Bruce 1984 |
| Diyari | Pama-Nyungan | OV | N+V | Austin 1981:162-164 |
| Dyirbal | Pama-Nyungan | OV | N+V | Anderson 1997 |
| Nez Perce | Plateau Penutian | OV | N+V | Aoki 1970:88-89 |
| Tunica | isolate | OV | N+V | Haas 1941:75 |
| Slave | Athabaskan | OV | N+V | Rice 1989:946-7 |
| Lakhota | Siouan-Catawba | OV | N+V | de Reuse 1994, Rood & Taylor 1996:454 |
| Koasati | Muskogean | OV | N+V | Kimball 1991:347, 469-70 |
| Damana | Chibchan | OV | N+V | Quesada 1999 |
| Imbabura Quechua | Quechuan | OV | N+V | Cole 1982:198 |
| Yoruba | NC: Benue-Congo | VO | N+V | Awoyale 1997 |
| Ewe | NC: Kwa | VO | N+V | Anderson 1997 |
| English* | Indo-European | VO | N+V | Anderson 1997 |
| Russian* | Indo-European | VO | N+V | Kiparsky 1975:345-350; JN |

| | | | | |
|------------|-----------------|----|-----|---|
| Finnish | Uralic | VO | N+V | Serebrennikov & Kert 1958:279-80 |
| Armenian | Indo-European | VO | N+V | Anderson 1997 |
| Vietnamese | Austroasiatic | VO | N+V | R. Rhodes, p.c. |
| Yagua | Zaparo-Yagua | VO | N+V | Anderson 1997 |
| Turkana | Eastern Nilotic | VO | V+N | Dimmendaal 1983:292-5 |
| Swahili | NC: Benue-Congo | VO | V+N | S. Mchombo, p.c. |
| Arabic | AA: Semitic | VO | V+N | Kaye 2005; A. Kaye, p.c. |
| Pulaar | NC: Atlantic | VO | V+N | Anderson 1997 |
| Hausa | AA: Chadic | VO | V+N | Anderson 1997 |
| Berber | AA: Berber | VO | V+N | Anderson 1997 |
| Irish | Indo-European | VO | V+N | Anderson 1997 |
| French | Indo-European | VO | V+N | NM |
| Mandarin | Sino-Tibetan | VO | V+N | Li & Thompson 1981:73ff., Shibatani 1990:241 |
| Tagalog | AN: WMP | VO | V+N | Zuraw 2006, Anderson 1997 |
| Samoan | AN: Polynesian | VO | V+N | Mosel & Hovdhaugen |
| Nisgha | Tsimshianic | VO | V+N | Tarpent 1999 |
| Tzutujil | Mayan | VO | V+N | Dayley 1984:190-191 |
| Cayuvava | isolate | VO | V+N | Key 1967:43-44 |
| Mapudungun | isolate | VO | V+N | Smeets 1989:423 |

| | | | | |
|----------------|--------------|------|-----|------------------|
| Bininj Gun-wok | Gunwingguan | None | N+V | Evans 2003:327-8 |
| Warlpiri | Pama-Nyungan | None | N+V | Nash 1986:37-8 |
| Yimas | Lower Sepik | None | N+V | Foley 1991 |

Appendix D: Languages with instrumental affixes

| Language | Family | Clause order | Affix order |
|-----------------|-------------------|-----------------|----------------|
| Achumawi | Palaihnihan | VO | Ins+V |
| Atsugewi | Palaihnihan | free | Ins+V |
| Biloxi | Siouan-Catawba | OV | Ins+V |
| Blackfoot | Algonquian | free | V+Ins |
| Chayahuita | Cahuapanan | OV | Ins+V |
| Comanche | Uto-Aztecan | OV | Ins+V |
| Cree | Algonquian | free | V+Ins |
| Crow | Siouan-Catawba | OV | Ins+V |
| Diegueño | Yuman | OV | Ins+V |
| Eastern Pomo | Pomoan | OV | Ins+V |
| Haida | isolate | OV | Ins+V |
| Ineseño Chumash | Chumashan | VO | Ins+V |
| Kawaiisu | Uto-Aztecan | OV | Ins+V |
| Klamath | Klamath-Sahaptian | free | Ins+V |
| Kutenai | isolate | VO | V+Ins |
| Lakhota | Siouan-Catawba | OV | Ins+V |
| Luißeño | Uto-Aztecan | VO | Ins+V |
| Lule | Lule-Vilela | OV | Ins+V |
| Maidu | Maidun (CPP) | OV | Ins+V |

| | | | |
|-----------|-------------------------|------|-------|
| Miwok | Miwok-Costanoan (CPP) | free | Ins+V |
| Nez Perce | Klamath-Sahaptian (CPP) | free | Ins+V |
| Ojibwa | Algonquian | free | V+Ins |
| Ponca | Siouan-Catawba | OV | Ins+V |
| Sahaptin | Klamath-Sahaptian (CPP) | VO | Ins+V |
| Shoshoni | Uto-Aztecan | OV | Ins+V |
| Tinrin | Oceanic (Austronesian) | VO | Ins+V |
| Tlingit | Na-Dene | OV | Ins+V |
| Tonkawa | Isolate | OV | Ins+V |
| Wappo | Isolate | OV | Ins+V |
| Yana | isolate | VO | Ins+V |

CPP = California and Plateau Penutian; Ins+V = instrumental affix precedes verb root;

V+Ins = instrumental affix follows verb root; data from Mithun (1999), Palancar (1999),

or Bickel and Nichols (2002ff), Adelaar & Muysken 2004 (Lule, Chayahuita), Osume

1995 (Tinrin), M. Dryer (Kutenai).

Appendix E. Frequencies across continents.

| | Africa, Eurasia | Pacific | North America | South America |
|--------------------|--------------------|---------|------------------|------------------|
| Productive NI | 5 | 13 | 18 | 3 |
| Unproductive | 1 | 3 | 4 | 0 |
| N + V | 5 | 10 | 17 | 4 |
| V + N | 1 | 6 | 4 | 1 |
| OV | 5 | 7 | 8 | 2 |
| VO | 1 | 7 | 12 | 3 |
| Productive & N+V | 4 | 7 | 14 | 2 |
| Productive & V+N | 1 | 6 | 4 | 1 |
| Unproductive & N+V | 1 | 3 | 4 | 0 |
| Unproductive & V+N | 0 | 0 | 0 | 0 |
| OV & N+V | 4 | 6 | 8 | 2 |
| OV & V+N | 1 | 1 | 0 | 0 |
| VO & N+V | 1 | 2 | 8 | 2 |
| VO & V+N | 0 | 5 | 4 | 1 |

References

- Abbott, Clifford. 2000. *Oneida*. Munich: Lincom Europa.
- Adelaar, Willem F. H., and Muysken, Pieter C. 2004. *The Languages of the Andes*.
Cambridge: Cambridge University Press.
- Adelman, Allison. 2002. On the status of N-V verbal compounds in Catalan. Senior Honors Thesis, Swarthmore College.
- Aissen, Judith. 2003. Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory* 21.435--483.
- Anderson, Stephen R. 1992. *A-morphous morphology*. Cambridge: Cambridge University Press.
- Anderson, Heather. 1997. The optimal compound: Optimality theory and the typology of synthetic compounds. *Optimal viewpoints*, ed. by Stuart Davis. Bloomington, IN: Indiana University Linguistics Club Publications.
- Aoki, Haruo. 1970. *Nez Perce Grammar*: UCPL 62. Berkeley-Los Angeles: University of California Press.
- Armbruster, Charles Hubert. 1960. *Dongolese Nubian: A grammar*. Cambridge: Cambridge University Press.
- Austin, Peter. 1981. *The grammar of Diyari, South Australia*. Cambridge: Cambridge University Press.
- Awoyale, Yiwola. 1997. Object positions in Yoruba. *Object positions in Benue-Kwa*, ed. by Rose-Marie Déchaine and Victor Manfredi, 7--30. The Hague: Holland Academic Graphics.

- 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. 1997. Complex predicates and agreement in polysynthetic languages.
Complex Predicates, ed. by Alex Alsina, Joan Bresnan, and Peter Sells, 247--288.
Stanford: CSLI.
- Baker, M., R. Aranovich, and L. A. Galluscio. 2005. Two types of syntactic noun
incorporation: Noun incorporation in Mapudungun and its typological
implications. *Language* 82.138--176.
- Bauer, Laurie. 2001. Compounding. In *Language Typology and Language Universals*,
eds. Martin Haspelmath, Ekkehard Koenig, Wulf Oesterreicher and Wolfgang
Raible, 695-707. Berlin-NY: Walter de Gruyter.
- Bickel, Balthasar, and Johanna Nichols. 2002ff. The Autotyp research program.
<http://www.uni-leipzig.de/~autotyp>
- Bickel, Balthasar, and Johanna Nichols. 2006. Oceania, the Pacific Rim, and the theory
of linguistic areas. Paper presented at the 32nd Annual Meeting of the Berkeley
Linguistics Society.
- Bickel, Balthasar, Dirk P. Janssen and Fernando Zúñiga. 2006. Randomization tests in
language typology. *Linguistic Typology* 10.419-440.
- Bossong, G. 1985. *Differentielle Objektmarkierung in den Neuiranischen Sprachen*.
Tübingen: Gunter Narr Verlag.
- Braine, Jean Critchfield. 1970. Nicobarese grammar (Car Dialect). Ph.D. dissertation,
University of California, Berkeley.

- Bricker, Victoria R. 1978. Antipassive constructions in Yucatec Maya, ed. by Nora C. England, 3-24. Columbia: University of Missouri-Columbia.
- Bruce, Les. 1984. *The Alambalak language of Papua New Guinea (East Sepik)*. (Pacific Linguistics C-81). Canberra: Australian National University.
- Bybee, Joan, William Pagliuca, and Revere Perkins. 1990. On the asymmetries in the affixation of grammatical material. *Studies in typology and diachrony: Papers presented to Joseph H. Greenberg on his 75th birthday*, ed. by William Croft, Keith Denning, and Suzanne Kemmer, 1--42. Amsterdam: Benjamins.
- Chirikba, Viacheslav A. 2003. *Abkhaz*. Munich: Lincom Europa.
- Chung, Sandra. 1978. *Case Marking and Grammatical Relations in Polynesian*. Austin: University of Texas Press.
- Cole, Peter. 1982. *Imbabura Quechua*. Amsterdam: North Holland.
- Dahl, Östen. 2004. *The growth and maintenance of linguistic complexity*. Amsterdam: Benjamins.
- Davidson, M. 2002. Studies in Southern Wakashan (Nootkan) grammar. Ph.D. dissertation, SUNY Buffalo.
- Dayley, Jon P. 1985. *Tzutujil grammar*. Berkeley, CA: University of California Press.
- Dayley, John P. 1989. *Tümpisa (Panamint) Shoshone grammar*. Berkeley, CA: University of California Press.
- de Reuse, Willem. 1994. Noun incorporation in Lakota (Siouan). *International Journal of American Linguistics* 60.199--260.
- Dimmendaal, Gerritt. 1983. *The Turkana Language*. Dordrecht-Cinnaminson: Foris.

- Dixon, R. M. W., and Alexandra Aikhenvald. 1999. *The Amazonian Languages*. Cambridge: Cambridge University Press.
- Donohue, Mark. 1999a. *A grammar of Tukang Besi*. Berlin: Mouton de Gruyter.
- Donohue, Mark. 1999b. *Warembori*. Munich: LINCOM Europa.
- Donohue, Mark. 2001. Split intransitivity and Saweru. *Oceanic Linguistics* 40:321-336.
- Dryer, Matthew. 1989. Large linguistic areas and language sampling. *Studies in Language* 13:257-292.
- Dryer, Matthew. 1992. The Greenbergian word order correlations. *Language* 68:81-138.
- Dryer, Matthew. 1997. On the 6-way word order typology. *Studies in Language* 21:69-103.
- Dunn, Michael John. 1999. A grammar of Chukchi. Ph.D. dissertation, Australian National University.
- Enrico, John. 2003. *Haida syntax*. Lincoln, NE: University of Nebraska Press.
- Enrique-Arias, Andres. 2002. Accounting for the position of verbal agreement morphology with psycholinguistic and diachronic explanatory factors. *Studies in Language* 26:1-32.
- Evans, Nicholas. 2003. *Bininj Gun-Wok: A pan-dialectal grammar of Mayali, Kunwinjku and Kune*. Canberra: Research School of Pacific and Asian Studies, Australian National University.
- Everett, Dan, and Barbara Kern. 1997. Wari': *The Pacaas Novos language of Western Brazil*. London: Routledge.

- Ezard, Bryan. 1992. Tawala derivational prefixes: A semantic perspective. *Papers in Austronesian Linguistics* 2, ed. Malcolm D. Ross, 147-250. (Pacific Linguistics A-82.) Canberra: The Australian National University.
- Foley, William. 1991. *The Yimas language of New Guinea*. Stanford: Stanford University Press.
- Galloway, Brent D. 1993. *A Grammar of Upriver Halkomelem*. Berkeley, CA: University of California Press.
- Gerdts, Donna B. 1998. Incorporation. In *The Handbook of Morphology*, eds. Andrew Spencer and Arnold M. Zwicky, 84-. London: Blackwell.
- Gerdts, Donna B. 2003. The morphosyntax of Halkomelem lexical suffixes. *International Journal of American Linguistics* 69:345-356.
- Gràcia, Luísa, and Olga Fullana. 1999. On Catalan verbal compounds. *International Journal of Latin and Romance Linguistics* 11.39--61.
- Green, Ian. 1989. *Marrithiyel: A language of the Daly River region of Australia's Northern Territory*. Ph.D. dissertation, Australian National University.
- Haas, Mary R. 1941. Tunica. *Handbook of American Indian Languages* 4, ed. by Franz Boas, 1--143.
- Haiman, John. 1980. *Hua: A Papuan language of the eastern highlands of New Guinea*. Amsterdam: Benjamins.
- Haude, Katharina. 2006. *A Grammar of Movima*. Ph.D. dissertation, Radboud University, Nijmegen.
- Heath, Jeffrey. 1978. *Ngandi grammar, texts, and dictionary*. Canberra: Australian Institute of Aboriginal Studies.

- Henderson, James. 1995. *Phonology and grammar of Yele*. Canberra: Research School of Pacific and Asian Studies, Australian National University.
- Houser, Michael J., and Maziar Toosarvandani. 2006. A nonsyntactic template for syntactic noun incorporation. Paper presented at the Annual Meeting of the Linguistic Society of America, Albuquerque, NM.
- Jacobsen, William H. Jr. 1964. A grammar of the Washo language. Ph.D. dissertation, University of California, Berkeley.
- Jacobsen, William H., Jr. 1980. Washo bipartite verb stems. In Kathryn Klar, Margaret H. Langdon, and Shirley Silver, eds., *American Indian and Indo-European studies: papers in honor of Madison S. Beeler*, 85-99. The Hague: Mouton.
- Jespersen, Otto. 1948. *Growth and Structure of the English Language*. Oxford: Blackwell.
- Karahashi, Fumi. 2000. *Sumerian compound verbs with body-part terms*. Ph.D. dissertation, University of Chicago.
- Kastovsky, Dieter. 2006. Typological changes in derivational morphology. In *The Handbook of the History of English*, eds. Ans van Kemenade and Bettelou Los, 151-176. Malden-Oxford-Carlton: Blackwell.
- Kaye, Alan S. 2005. Semantic transparency and number marking in Arabic and other languages. *Journal of Semitic Studies* 1.153--196.
- Kayne, Richard. 1994. *The antisymmetry of syntax*. Cambridge: Cambridge University Press.
- Key, Harold H. 1967. *Morphology of Cayuvava*. The Hague-Paris: Mouton.
- Kimball, Geoffrey D. 1991. *Koasati Grammar*. Lincoln: University of Nevada Press.

- Kiparsky, Valentin. 1975. *Russische historische Grammatik, Band III: Entwicklung des Wortschatzes*. Heidelberg: Carl Winter.
- Klingebiel, Kathryn. 1989. *Noun + Verb Compounding in Western Romance*. (UCPL 113.) Berkeley-Los Angeles: University of California Press.
- Kozinsky, Ishaak. 1981. Nekotorye grammaticheskie universalii v podsystemax vyrazhenija subjektno-objektnyx otnoshenij. Ph.D. dissertation, Moscow State University.
- Kroeker, Menno. 2001. A descriptive grammar of Nambikuara. *International Journal of American Linguistics* 67.1--87.
- Leer, Jeff. 1977. Introduction. In Lawrence, Erma, ed. *Haida Dictionary*. Fairbanks: ANLC.
- Lehmann, Winfred P. Proto-Indo-European compounds in relation to other Proto-Indo-European syntactic patterns. *Acta Linguistica Hafniensia* 12.1-20.
- Lehmann, Winfred P. 1975. A discussion of compound and word order. In Charles N. Li, ed., *Word Order and Word Order Change*, 149-162. Austin: University of Texas Press.
- Li, Charles N., and Thompson, Sandra A. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley, CA: University of California Press.
- Malouf, Robert. 1999. West Greenlandic noun incorporation in a monohierarchical theory of grammar. Gert Webelhuth, Andreas Kathol, and Jean-Pierre Koenig, eds., *Lexical and Constructional Aspects of Linguistic Explanation*, 47-62. Stanford: CSLI Publications.

- Marchand, Hans. 1969. *The Categories and Types of Present-Day English Word Formation*. Munich: Beck.
- Mardirussian, Galust. 1975. Noun incorporation in universal grammar. *CLS* 11.383-389.
- Maslova, Elena. 2000. A dynamic approach to the verification of distributional universals. *Linguistic Typology* 4.307-333.
- Massam, Diane. 2001. Pseudo incorporation in Niuean. *Natural Language and Linguistic Theory* 19:153-197.
- Mattissen, Johanna. 2003. *Dependent-Head Synthesis in Nivkh: A contribution to a typology of polysynthesis*. Amsterdam/Philadelphia: Benjamins.
- Melnar, Lynette R. 2004. *Caddo verb morphology*. Studies in the native languages of the Americas. Lincoln, NE: University of Nebraska Press.
- Miner, Kenneth L. 1986. Noun stripping and loose incorporation in Zuni. *International Journal of American Linguistics* 52.242--254.
- Mithun, Marianne. 1984. The evolution of noun incorporation. *Language* 60.847--894.
- Mithun, Marianne. 1999. *The Languages of Native North America*. Cambridge: Cambridge University Press.
- Morgan, Lawrence. 1991. A description of the Kutenai language. Ph.D. dissertation, University of California, Berkeley.
- Mosel, Ulrike, and Even Hovdhaugen. 1992. *Samoan Reference Grammar*. Oslo: Scandinavia University Press.
- Nakayama, T. 1997. Discourse-pragmatic dynamism in Nun-chah-nulth (Nootka) morphosyntax. Ph.D. dissertation, University of California, Santa Barbara.

- Newman, P. 2000. *The Hausa language: An encyclopedic reference grammar*. New Haven, CT: Yale University Press.
- Newmeyer, Frederick J. 2000. On the reconstruction of 'Proto-World' word order. In Chris Knight, Michael Studdert-Kennedy, and James R. Hurford, eds., *The Evolutionary Emergence of Language: Social function and the origins of linguistic form*, 372-389. Cambridge: Cambridge University Press.
- Nichols, Michael J. P. 1974. Northern Paiute historical grammar. PhD dissertation, University of California, Berkeley.
- Osborne, C. R. 1974. *The Tiwi language*. Canberra: AIAS.
- Osumi, Midori. 1995. *Tinrin Grammar*: Oceanic Linguistics Special Publication no. 25. Honolulu: University of Hawaii Press.
- Palancar, Enrique L. 1999. Instrumental prefixes in Amerindian languages: An overview to their meanings, origin, and functions. *STUF* 52(2).151-166.
- Plank, Frans, Thomas Mayer, Tatsiana Mayorava and Elena Filimonova. n.d. The Universals Archive. <http://typo.uni-konstanz.de/archive/>.
- Progovac, Ljiljana. 2007. Layering of grammar: Vestiges of protosyntax in modern-day languages. Geoffrey Sampson, ed., *Language Complexity as an Evolving Variable*, 203-212. Oxford: Oxford University Press.
- Pycha, Anne. 2008. Morphological sources of phonological length. Ph.D. dissertation, UC Berkeley.
- Quesada, J. Diego. 1999. Chibchan, with special reference to participant highlighting. *Linguistic Typology* 3-2.209-258.

- Rehg, Kenneth L. 1981. *Ponapean reference grammar*. Honolulu, HI: University Press of Hawaii.
- Reid, Nicholas. 2003. Phrasal verb to synthetic verb: Recorded morphosyntactic change in Ngan'gityemerri. In *The Non-Pama-Nyungan Languages of Northern Australia*, ed. Nicholas Evans, 95-123. Canberra: Research School of Pacific and Asian Studies, Australian National University.
- Rice, Keren. 1989. *A Grammar of Slave*. Berlin: Mouton de Gruyter.
- Robertson, John. 1980. *The structure of pronoun incorporation in the Mayan verbal complex*. New York: Garland.
- Rood, David S., and Taylor, Allan R. 1996. Sketch of Lakhota, a Siouan language. In *Handbook of North American Indians: Languages*, ed. Ives Goddard, 440-482. Washington: Smithsonian Institution.
- Rosen, Sara. 1989. Two types of noun incorporation: A lexical analysis. *Language* 65.294-317.
- Runner, Jeffrey T., and Aranovich, Raul. 2003. Noun incorporation and rule interaction in the lexicon. In *Proceedings of the 10th International Conference on Head-Driven Phrase Structure Grammar*, ed. Stefan Mueller, 359-379. Stanford: CSLI Publications. <http://csli-publications.stanford.edu/HPSG/4/> [accessed October 2007]
- Sadock, Jerrold M. 1980. Noun incorporation in Greenlandic. *Language* 56.300--319.
- Sadock, Jerrold M. 1986. Some notes on noun incorporation. *Language* 62.19--31.
- Saeed, John I. 1999. *Somali*. Amsterdam-Philadelphia: Benjamins.
- Saltarelli, M. 1988. *Basque*. Croom Helm: London.

- Sapir, Edward. 1910. *Yana texts* (UCPAAE 9:1). Berkeley, CA: University of California Press.
- Sapir, Edward. 1911. The problem of incorporation in American languages. *American Anthropologist* 13.250-282.
- Sapir, Edward. 1922. The Takelma languages of southwestern Oregon. *Handbook of American Indian Languages* (BAE Bulletin 40), ed. by Franz Boas, 1--296.
- Serebrennikov, B. A., and G. M. Kert. 1958. *Grammatika finskogo jazyka*. Moscow-Leningrad: AN SSSR.
- Shibatani, Masayoshi. 1990. *The languages of Japan*. Cambridge: Cambridge University Press.
- Siewierska, Anna and Dik Bakker. 1996. The distribution of subject and object agreement and word order type. *Studies in Language* 20.1-20.
- Singh, Naorem Saratchandra. 2006. *A Grammar of Paite*. New Delhi: Mittal Publications.
- Smeets, Ineke. 1989. *A Mapuche Grammar*. Leiden: Rijksuniversiteit.
- Soe, Myint. 1999. A Grammar of Burmese. Ph.D. Dissertation, University of Oregon.
- Spencer, Andrew. 1995. Incorporation in Chukchi. *Language* 71.439--89.
- Stonham, John T. 2004. *Linguistic theory and complex words: Nuuchahnulth word formation*. New York: Palgrave Macmillan.
- Strom, Clay. 1992. *Retuarã Syntax*. (Studies in the Languages of Colombia, 3.) Arlington, TX: SIL and University of Texas-Arlington.
- Stump, Gregory T. 1992. On the theoretical status of position class restrictions on inflectional affixes. *Yearbook of morphology*, ed. by G. Booij and J. van Marle, 211-41. Dordrecht: Kluwer.

- Suttles, Wayne. 2004. *Musqueam reference grammar*. Vancouver: University of British Columbia Press.
- Tarpent, Marie-Lucie. 1987. A grammar of the Nisgha language. Ph.D. dissertation, University of Victoria.
- Tosco, Mauro. 2004. Between zero and nothing: Transitivity and noun incorporation in Somali. *Studies in Language* 28:83-104.
- van Geenhoven, Veerle. 1998. *Semantic Incorporation and Indefinite Descriptions: Semantic and syntactic aspects of noun incorporation in West Greenlandic*. Stanford: CSLI.
- Velazquez-Castillo, Maura. 1996. *The grammar of possession: Inalienability, incorporation and possessor ascension in Guaraní*. Amsterdam: John Benjamins.
- Viegas Barros, J. Pedro. 2001. Evidence for a Lule-Vilela genetic relationship. *Liames* 1.107--126.
- Walsh, Michael. 1987. The impersonal verb construction. *Language topics: Essays in honour of Michael Halliday*, Vol. 1, ed. by R. Steele and T. Threadgold, 425--438. Amsterdam: Benjamins.
- Watkins, Laurel J., and McKenzie, Parker. 1984. *A grammar of Kiowa*. Lincoln, NE: University of Nebraska Press.
- Williams, Edwin. 1981. On the notions 'lexically related' and 'head of a word'. *Linguistic Inquiry* 12.245--274.
- Wolfart, H. C. 1973. *Plains Cree: A grammatical study*. Philadelphia: American Philosophical Society.

Zide, Arlene K. 1997. Pronominal and nominal incorporation in Gorum. *Languages of tribal and indigenous peoples of India: The ethnic space*, ed. by Anvita Abbi, pp. Delhi: Motilal Banarsidass Publishers.

Languages with noun incorporation (N = 219)

