Harley, Heidi & Harvey, Meg. 2025. Plural reference dominance, markedness and semantic categorization in Hiaki pluralia tantum. *Glossa: a journal of general linguistics* 10(1). pp. 1–39. DOI: https://doi.org/10.16995/glossa.16308

Open Library of Humanities

Plural reference dominance, markedness and semantic categorization in Hiaki pluralia tantum

Heidi Harley, University of Arizona, hharley@arizona.edu

Meg Harvey, William and Mary, maharvey01@wm.edu

We provide a description and analysis of "pluralia tantum" (PT) nouns in the Southern Uto-Aztecan language Hiaki (Yaqui, Yoeme, YAQ ISO 639-3). We find that these nouns, which require plural morphosyntactic marking regardless of notional number, fall into several semantic categories. We then model the behavior of number marking in Hiaki using a Distributed Morphology framework. We analyze apparent mismatches in the agreement system that prima facie appear problematic for Corbett's (2019) Agreement Hierarchy. We propose that they result from a distinction between purely morphological 'Concord' features on the noun that can be independent from semantically conditioned 'Index' features, taking the Concord/Index distinction from Wechsler and Zlatić (2000; 2003). Index features determine choice of suppletive verbal form, while Concord features control nominal number marking, adjectival and determiner number marking, and the form of anaphoric pronominals. The conclusion is that number-conditioned verbal suppletion is distinct from true verbal agreement. We conclude by discussing whether a frequentist account of the emergence of individual PT nouns might apply in the Hiaki case, i.e. whether plural-reference dominance in these semantic categories might have driven grammaticization of the nominal as a PT noun, and argue against this possibility.

1 Introduction

Hiaki (Yaqui, Yoeme, YAQ ISO 639-3) is a Southern Uto-Aztecan language of the Taracahitan subfamily, spoken in Sonora, Mexico and in the southwestern United States. Previous documentation has remarked that certain nouns in the language exhibit 'pluralia tantum' (PT) behavior, requiring plural morphosyntactic marking even when used with singular reference.

In this paper, we first document the robust nature of the PT category in Hiaki (atypical for the family). We begin by laying out a definition of pluralia tantum situated in Corbett's typological treatment, grounded in his hypothesized Agreement Hierarchy. We then situate Hiaki PT behavior in this context. In §2, we describe and categorize the data set resulting from our investigation of Hiaki PT nouns in extant lexicographic documentation. In §3, we take a detailed look at the categories of Hiaki PT nouns, concluding that PT membership is connected to particular semantic categories, and argue against previous suggestions that PT membership for borrowed Spanish nouns in Hiaki is a reflection of a nativization process for borrowed nouns. In §4, we propose a formal model of pluralia tantum morphosyntax in Hiaki using the Distributed Morphology framework. We exploit Wechsler and Zlatić (2000; 2003)'s distinction between 'Index' features and 'Concord' features to account for the differing behavior of nominal concord and anaphoric agreement on the one hand and suppletive verbal agreement on the other. We explore the implications for the Agreement Hierarchy of Corbett (1979; 2003; 2006; 2013; 2019; 2022; 2023), arguing that the suppletive verbal agreement facts do not pose a challenge to the Agreement Hierarchy. Section 5 explores the implications for theories of grammaticization and morphological markedness, looking at plural-reference dominance for translation equivalents of some Hiaki PT nouns, following the methodology of Haspelmath and Karjus (2017). Our overall conclusion is that pluralia tantum categories in Hiaki line up well with the conceptual categories that form the basis for Grimm (2018)'s Scale of Individuation, but are hard to reconcile with a frequentist account of their diachronic development.

A 'plurale tantum' noun is a noun that necessarily exhibits plural grammatical properties, even when used with unit reference. These properties might be morphological or morphosyntactic, or both. In Corbett (2019)'s canonical typology approach, a form like Russian *sani*, 'sledge', is a canonical example of a plurale tantum noun, because it exhibits the behavior expected of plurals through the whole range of contexts described in his Agreement Hierarchy, while at the same time composing directly with numeral 'one', exhibiting unit reference: *odn-i san-i* one-PL sledge- PL 'one sledge'. Russian *sani* stands in contrast to Tsez *xexbi*, which shows the morphological characteristics of plural inflection locally but fails to participate in any of the morphosyntactic agreement phenomena that a plural noun would otherwise normally exhibit. In a canonical typology view of pluralia tantum, *sani* and *xexbi* represent opposite ends of a spectrum, as represented in **Table 1**.[1]

[1] The typology of English pluralia tantum nouns like *pliers* or *pants* is somewhat atypical, since the morphosyntax of the English noun phrase forbids unit reference without a singularizing classifier: **one pliers* vs *one pair of pliers*, an analytical challenge we do not address in this paper. In other respects, however, especially in their semantic clustering behavior, English does instantiate properties of pluralia tantum nouns with which we are engaged here, and we will return to some English examples below.

In unit-reference contexts, this noun…	…bears plural morphology	…triggers plural concord	…triggers plural verb agreement	…antecedes only plural anaphors
Russian *sani*	y	y	y	y
Tsez *xexbi*	y	n	n	n

Table 1: Exemplars of most and least canonical pluralia tantum behavior.

In the Hiaki language, pluralia tantum nouns behave like the canonical Russian case, with one significant exception. They bear plural morphology (1a), they trigger plural concord within the DP (1b), and they antecede only plural anaphors (1c, d), even when they have unit reference.

(1) a. *livrom*[2] **livro*
livro-m livro
book-PL book
"book(s)"

henom **heno*
heno-m heno
shoulder-PL shoulder
"shoulder(s)"

b. *ume* *livrom*
u-me livro-m
DET-PL book-PL
"the/a book(s)"

**uu* *livro(m)*
uu livro-(m)
DET.SG book-(PL)
"the/a book(s)"

ume *henom*
u-me heno-m
DET-PL shoulder-PL
"the/a shoulder(s)"

**uu* *heno(m)*
uu heno-(m)
DET.SG shoulder-(PL)
"the/a shoulder(s)"

c. *Inepo livrom hinuk. Im am teeka.*
Inepo livro-m hinu-k im am/*aa=teeka.
1SG.NOM book-PL buy-PFV here 3PL.ACC/*3SG.ACC=lay
"I bought a book. (I) laid it here." (Lit: "(I) laid them here.")

[2] We provide our Hiaki data in the English-based orthography of the Pasqua Yaqui tribe, so e.g., 'h', 'v', 'bw' correspond to 'j', 'b' and 'bu' in the Spanish-based orthography in Sonora.

d. *Henompo ne wante. Uu chuu'u am keeka*
heno-m-po = ne wante uu chuu'u am/*aa = keeka
shoulder-PL-in = 1SG.NOM hurt the.NOM dog.NOM 3PL.ACC/*3SG.ACC = bite.PFV
"My shoulder hurts. The dog bit it." (Lit: "The dog bit them.")

The inflectional behavior of PT nouns contrasts with the inflectional behavior of 'regular' count nouns, which show a mandatory semantically conditioned singular/plural inflectional contrast (2a, b) which triggers mandatory agreement on the determiner (2c, d).

(2) a. *Nee hiokareota hinuk.*
nee hiokareo-ta hinu-k
1.SG.NOM pencil-ACC.SG buy-PFV
"I bought a pencil." (Cannot mean: "I bought pencils")

b. *Nee hiokareom hinuk.*
nee hiokareo-m hinu-k
1.SG.NOM pencil-PL buy-PFV
"I bought pencils." (Cannot mean: "I bought a pencil")

c. *Nee hunuka hiokoreota hinuk.*
nee hunuka hiokoreo-ta hinu-k
1.SG.NOM that.ACC.SG pencil-ACC.SG buy-PFV
"I bought that pencil." (Cannot mean: "I bought those pencils")

d. *Nee hunume hiokareom hinuk.*
nee hunu-me hiokareo-m hinu-k
1.SG.NOM DET-PL pencil-PL buy-PFV
"I bought those pencils." (Cannot mean: "I bought that pencil")

The one exception to the otherwise uniformly 'plural' morphosyntax of pluralia tantum nouns in Hiaki involves verbal suppletion conditioned by number. Pluralia tantum nouns with unit reference require singular forms of verbs (3a), and only PT nouns with true plural reference co-occur with plural forms of verbs (3b).

(3) a. *Hunume livrom ama mesapo vooka.*
hunu-me livro-m ama mesa-po vo'oka
DET-PL book-PL there table-on lie.SG
"That book is lying on the table."

b. *Hunume livrom ama mesapo to'oka.*
hunu-me livro-m ama mesa-po to'oka
DET-PL book-PL there table-on lie.PL
"Those books are lying on the table."

There are only a few verbs which show number agreement. Most verbs in Hiaki do not agree in number, and the few verbs that do indicate argument number always do so via suppletion, as in

(3) above, a point to which we will return below. Prima facie, however, the failure of Hiaki PT nouns to trigger verbal agreement, despite requiring plural anaphora, constitutes a violation of Corbett (2019)'s Agreement Hierarchy, since verbal agreement is lower on the hierarchy than anaphoric agreement—anaphoric agreement in a system is predicted to entail verbal agreement in that system, according to the AH typology. We will argue below (section 6.1) that in fact the AH typology is not violated by Hiaki. Instead, we should take this as supporting evidence for the idea that suppletive plural agreement isn't 'true' syntactic agreement, following Harley et al. (2016) and Bobaljik & Harley (2017).[3]

Next we provide a more in-depth presentation of pluralia tantum nominal behavior, contrasting it with 'regular' number-marking nominal behavior.

2 Hiaki pluralia tantum

Most Hiaki nouns are countable (like *hiokareo* 'pencil' in (2) above), and mandatorily make a morphosyntactic number contrast between singular and plural. Singular nominals are unmarked in the nominative and bear a *-ta* case suffix in accusative and oblique contexts. Plural nominals bear the plural suffix *-(i)m,* which does not distinguish case. We again illustrate this in (4a) below, with a different count noun, *kuta* 'stick'. However, many nouns are pluralia tantum, including most mass nouns (4b) and many idiosyncratic count nouns (4c, d, e).

(4)		Sg	Pl	
	a.	kuta(ta)	kutam	"stick(-ACC), sticks"
	b.	*	vaa'am	"water"
	c.	*	supem	"dress(es)"
	d.	*	sutum	"fingernail(s)"
	e.	*	veho'orim	"lizard(s)"

2.1 Morphosyntax of PT nouns

All nouns that head an NP must be marked for number or case, as shown for the object noun *kuta* 'stick' in (5a)—in object position, it cannot occur without either a plural marker or the singular accusative case suffix. When a noun is incorporated (5b) or is the left-hand member of a compound (5c), it cannot be marked for number or case (Haugen & Harley 2012).

(5) a. *Aapo kutam/kutata/*kuta aman siuta.*
aapo kuta-**m**/kuta-**ta**/*kuta aman siuta
3SG stick-**PL**/stick-**ACC.SG**/stick there tear.TR
"S/he is splitting sticks/a stick over there."

[3] Versions of this conclusion have been proposed for similar phenomena in other languages elsewhere; see especially Durie (1986); for a typological perspective on number-conditioned suppletion in general, see Corbett (2000: Ch.8). For a theoretical approach to number-conditioned verbal suppletion in Kiowa, see Harbour (2008: Ch. 4). For a different take on the Hiaki data, see Thornton (2019: 538).

b. *Aapo aman kutasiute/*kutamsiute/*kutatasiute.*
aapo aman kuta-siute /*kuta-**m**-siute/*kuta-**ta**-suite
3SG there stick-tear.INTR/stick-**PL**-tear.INTR/STICK-**ACC.SG**-tear.INTR
"S/he is splitting wood over there"

c. *kuta wikui /*kutam wikui(m) /*kutata wikui(ta)*
kuta wikui/*kuta-m wikui(-m)/*kuta-ta wikui(-ta)/
stick lizard/stick-PL lizard(-PL)/stick-ACC.SG lizard(-ACC.SG)
'iguana'

This pattern holds for PT nouns as well. They must bear the plural number sufix when they head a separate NP (6a, c), and must lose their number inflection when incorporated (6b) or compounded (6d).

(6) a. *Inepo paanim/*paan Mariatau woi pesopo hinuk.*
inepo paan-im/*paan Maria-ta-u woi peso-po hinu-k
1SG.NOM bread-PL/*bread Maria-ACC-from two peso-for buy-PFV
"I bought the bread from Maria for two pesos."

b. *pan-hoa*
"to bread-make" > *panim* "bread"

c. *In mamam/*mam elesiiikile.*
in mama-M/*mam elesiiki-le
1SG.GEN hand-PL/*hand itchy-find
"My hand(s) are/is itchy"

d. *mam-pusim*
"hand-eyes" = "fingers" > *mamam* "hands/hand"

As we have seen in (2), in Hiaki, nominal number is marked on the head noun and on the determiner/demonstrative. As we saw above (3), and illustrate again in (7), number marking on countable pluralia tantum nouns in Hiaki is purely formal: Pluralia tantum forms with numeral 'one' are fine, (7a), in contrast to the incompatibility of 'one' with plural forms of regular count nouns, which have a plural entailment (7b). It is also worth noting that pluralia tantum concord appears on postposed adjectives and numerals (7c), and even on stranded adjectives and numerals in N-ellipsis constructions (7d).

(7) a. *Nee*[4] *wepul supem hinuk.*
nee wepul supe-m hinu-k
1SG.NOM one dress-PL buy-PFV
"I bought one dress."

[4] The pronominals *nee* and *inepo* are interchangeable first person singular nominative forms.

b. *Nee wepul hiokareota/*hiokareom hinuk.*
nee wepul hiokareo-ta/*hiokareo-m hinu-k
1.SG.NOM one writing.instrument-ACC.SG/*writing.instrument-PL buy-PFV
"I bought one pencil."

c. *Vanteam wikwak, tosaim.*
vantea-m wikwa-k tosai-m
flag-PL pull-PFV white-PL
"[They] hoisted the white flag."

d. *Uuchi wepulaim nee ya'aria.*
uuchi wepulai-m nee ya'a-ria
again one-PL 1SG.ACC make-APPL
"Make me one [tortilla] again." Elided: *tahkaim* 'tortilla(s)', a PT noun

This "one [PT noun]" construction is, so far as we know, acceptable for all pluralia tantum count nouns in Hiaki, including for paired body part nouns such as *puusim* 'eye(s)' and *wokim* 'leg(s)'.

2.2 Collecting and categorizing Hiaki PT nouns

In previous documentation, Hiaki PT have not received much attention, although they are often remarked on in passing. It is an old pattern in Hiaki. The earliest extant documentation, the early 17th-century Jesuit grammar later reprinted as Buelna (1890), notes (p. 44) the existence of nouns 'caracen de singular', i.e. 'lacking singular', and list several examples familiar from the modern language, e.g. *supem* 'el vestido' [dress(es)], *puusim* 'los ojos' [eye(s)], *nacam* 'las orejas' [ear(s)], as well as a couple that are no longer pluralia tantum, e.g. *vatatzim* 'la rana' [frog(s)], and *tzoquim* 'las estrellas' [star(s)].

The only suggestion of a generalization concerning this grammatical class that we have seen in previous literature involves the frequent application of the plural suffix to borrowed words. Molina et al. (1999: *-m*), list one sense of the plural suffix *-m* as a marker of borrowed Spanish terms:

> "3. marker of terms introduced by Spanish culture (especially tools; ex. *livrom*,
> 'book/books'; *martiom*, 'hammer/hammers'; *tisiriam*, 'scissors')"

In a similar vein, Estrada Fernández and Guerrero (2007: 3) suggest that Spanish bare nouns are generally borrowed with a plural or collective interpretation and hence receive the *-m* suffix.[5] It

[5] Estrada Fernández and Guerrero (2007) hypothesize that this may explain why Hiaki borrows some Spanish nouns in a form which includes the Spanish plural suffix as singular terms (e.g. *wakas* 'cow' from *vaca-s*); this also, however, is not a general property of Spanish borrowings, consider *kava'i* from *caballo* 'horse', or *chiiva* from *chiva* 'goat'.

certainly is the case that many Spanish borrowings have become pluralia tantum nouns in Hiaki (8a–g). However, many have not (e.g. 8h–n):

(8)

	Sg	Pl	
a.	*	*livrom*	‘book(s)’
b.	*	*leentem*	‘glasses’
c.	*	*waantem*	‘glove(s)’
d.	*	*saweam*	‘pants/shorts’
e.	*	*supem*	‘blouse/dress(es)’
f.	*	*laapisim*	‘pencil(s)’
g.	*	*mache’etam*	‘machete(s)’
h.	*chiiva*	*chiivam*	‘goat, goats’
i.	*kava’i*	*kava’im*	‘horse, horses’
j.	*laaven*	*lavenim*	‘violin, violins’
k.	*wakas*	*wakasim*	‘cow, cows’
l.	*pipa*	*pipam*	‘pipe, pipes’
m.	*mansaana*	*mansaanam*	‘apple, apples’
n.	*kaaro*	*karom*	‘car, cars’

We return to discussion of the PT status of borrowed nouns in section 3.1 below.

We undertook the first systematic review of PT nouns in the language by examining all the nouns in the Molina et al. (1999) dictionary, supplemented with forms and examples from the Estrada Fernández et al. (2004) dictionary. Of the 1421 noun headwords in the Molina et al. dictionary, 316 are given with the plural suffix. In other words, 22% of the headwords in the dictionary seem to be pluralia tantum, in contrast to previous assessments that suggested “there are only a few inherently plural nouns” (Dedrick & Casad 1999: 131). For a complete list, see the DOI linked under “Data Availability”.

While our starting point was identifying these headwords, where possible we also looked at existing example sentences using these terms and at times elicited new ones. Occasionally, we found a difference between dictionary example sentences and elicited speech. For example, both dictionaries only listed the PT form *veho’orim* ‘lizard(s)’. Estrada Fernandez et al. (2004) give the following example, in which *veho’orim* is marked for plural but receives a singular translation:

(9) *Ili uusi veho’orim hu’upa nawapo yeu wiikek.*
ili uusi veho’ori-**m** hu’upa nawa-po yeu wiike-k
little child lizard-**PL** tree root-at out pull-PFV
“The child pulled **a lizard** out of the root of the tree.”
Original Spanish translation: “El niño sacó **una lagartija** de la raiz del arbol.”
(Estrada Fernández et al. 2004: 63)[6]

[6] We have used the Arizona Hiaki orthography here for consistency, but the original example is presented in the Spanish-based Sonora Hiaki orthography: “*Ili uusi bejo’orim ju’upa nawapo yeu wiikek.*”

However, one speaker of Arizona Hiaki identified *veho'ori* 'lizard' as an acceptable singular form. We conclude that there can be lexical variation between speakers concerning the PT status of particular nouns.

Given the standard observation that pluralia tantum nouns are often drawn from recognizable semantic categories (see, e.g. Williams 1994: 12), we investigated whether we could make any plausible generalizations about which categories of referents are marked as pluralia tantum in Hiaki. This process yielded immediate and robust results, with almost all Hiaki PT nouns falling into one of several clear semantic categories.[7] We present the categories we found in (10), in an order suggested by Grimm (2018: 543)'s scalar organization of entity types exhibiting collective/singulative marking in four unrelated languages.

(10) **i. Liquids**

suerom, IV fluid; *vahkom,* lake, pond; *vaa'am,* water; *choomim,* phlegm; *mumum,* honey; *pi'ikim,* milk; *oppoam,* tears

ii. Substances

chiktitam, chewing gum; *chu'ukam,* resin, pitch; *techuniam,* grime, filth; *maatum,* charcoal; *raahum,* caked dried earth; *saavum,* soap; *tetamatum,* coal; *haakam,* mucus

iii. Foodstuffs

ainam, flour*; sopi'ichim,* overripe fruit*; miisam,* Eucharist wafer*; keesum,* cheese*; tahkaim,* tortilla*; sito'im,* jelly*; muunim,* beans*; ruenasim,* peaches*; nohim,* tamales; *saakim,* parched corn; *peonasim,* peas*; heseim,* brown tepary beans*; luusem,* candy, sweets*; gayeetom,* cookies

iv. Aggregates

teeham, hail; *tutukam,* scree, gravel; *ouvam,* coals, embers; *tiikom,* wheat; *kovalam,* sewing pins; *pahtiam,* aspirin; *chuhtiam,* wood chips, cloth scraps; *aulim,* clams

v. Fabric

hekam, canopy; *hiniam,* shawl; *kolcham,* quilt; *kamisetam,* undershirt; *kortiinam,* curtains; *loonam,* canvas, tarp; *movektiam,* ceremonial head cloth; *sekawam,* matachin crown ribbon; *vanteam,* flag; *wanwoochim,* burlap; *chaketonim,* coat, jacket; *hipetam,* bed, mat

vi. Foliage/ Leafy plants

mamyam, certain greens; *paakam,* hay; *chichiham,* mistletoe; *chunahkam,* mesquite flower; *avi'itom,* lambs' quarters; *nakkaim,* Santa Rita prickly pear; *ko'apa'im,* plant for snake bite; *mavem,* a kind of plant

vii. Bichos, creepy-crawlies (cf. vermin)

veho'orim, lizard; *chinchim,* chiggers, bedbugs; *eesukim,* sugar ants; *poowim,* newt, salamander; *kuurum,* sand fleas, *etem,* lice, fleas, *surem,* ancestors

[7] A reviewer asks whether the converse can also be done. That is, whether we can look at categories of words that are not PT and discuss them. Such a mirror image categorization project would require looking at ~80% of Hiaki nouns, however, what we can say impressionistically is that no clear groupings stand out in this inverse category.

viii. Body parts

pempe'im, heel; *wepe'im,* hips; *puusim,* eye; *wokim,* leg; *moe'esom,* tonsil; *tonom,* knee; *tamim,* teeth; *voam,* fur, feathers; *hemaha'achim,* lungs; *sana'im,* rib; *sutum,* nail, claw; *vi'am,* nape; *chomim,* anus; *chumim,* vagina; *choam,* crown of head; *choonim,* head hair, scalp; *himsim,* mustache; *tero'okim*, ankle

ix. Hand tools

chiivam, crowbar*; hilukiam*, musical rasper*; mache'etam,* machete*; paalam,* shovel*; plancham,* iron, *tepuam,* axe*; hoosom,* sickle*; nava'asom,* pocket knife; *sena'asom,* pascola's disk rattle; *liimam,* file; *ayam,* deer dancer rattle; *ehpam,* sword; *hi'ikiam,* needle; *lansam,* spear; *mule'etam,* crutches; *tetam,* gavel; *yaavem,* key; *martiom,* hammer

x. Multipart object

seve'im, fringe; *kookam,* necklace; *kananam,* bandolier; *tekuriam,* knob on antler; *ehkaleam,* ladder; *kartam,* mail; *livrom,* book; *tenevoim,* cocoon leg rattles; *tatakalim,* 'any forked object'; *luusim,* tail lights on car; *koyoolim,* jingle bells, bell belt; *pinsam,* tweezers; *chaptiam,* scissors[8]

xi. Disease

taakam, pustule; *chupuwaim,* little joint pains; *namuwam,* cataracts; *tomtiam,* smallpox; *huttiam,* skin rash; *he'oktiam,* hiccups; *huva'asam,* venereal disease; *sarampionim,* measles

xii. Constellations

Kaarom, the Big Dipper; *Vahtekoim,* the Pleiades; *Choki Araum,* the Big Dipper; *Napo Hisa'im* the Milky Way

xiii. Words/music

soonim, pascola dance tune; *kanariom,* first tune played; *limohnaim,* devotional song; *kavansam,* certain dance tune; *kuaktiteam,* naming ritual for infant; *team,* name; *kavayom,* notice, notification; *vihtam,* movies; *alavansam,* hymn

xiv. Abstract

rupaktiam, flames; *tenkuim,* dreams; *yeetem,* drowsiness; *eerim,* thoughts; *koloorim,* color; *kuhteerim,* anger; *huneewam,* knowledge, wisdom, *unum-po* at one o'clock

xv. Landscape

takalaim, certain enchanted mountain*; maayom* Mayo country; *hiakim,* Hiaki country; also six of the eight pueblos: *Bacum, Torim, Rahum, Potam, Belem, Vicam*

xvi. Groups

peloteam, baseball team; *kooparíam,* singer society; *ya'uram,* government, *morom,* moor society; *pasioneom,* participants in pahko; *hiponreom,* band (music); *tenanchim,* female litter bearers; *fariseom,* fariseo society; *hurasim,* chapayekas; *wo'orim,* twins

xvii. Categorization unclear

chaatim, shot, injection (hand tool?); *kanteelam,* candle (hand tool?)*; lakim,* lock (multipart object?)*; katom,* wooden/bone ball for shinny (???)*; bombam,* bomb (fireworks/stars/multipart object?)*; huham,* fart (abstract, nebulous?)*; tapehtim,* cane platform (woven – fabric?)

---

[8] Both *chaptiam* and *pinsam* also might belong in the 'hand tools' category.

It is important to note that the connection between being a PT noun and denoting in a relevant semantic category is a one-way implication: If a Hiaki noun is a PT noun, it belongs to one of these semantic categories. However, there are nouns that denote in these categories which are not grammatically PT. Consider, e.g. *kafe*, 'coffee' and *ohvo,* 'blood' which denote liquids but are formally singular, or *choki tachiria* 'starlight', which denotes an extensionally cumulative concept, but is formally singular. Such cases play an important role in the featural analysis we propose in section 4.2 below.

2.3 Hiaki PT nouns in a cross-linguistic context

We see in this categorization of Hiaki PT nouns some similarities with PT behavior in other languages. Koptjevskaya-Tamm & Wälchli (2001: 630), for example, note that substances, complex artifacts (i.e., "multipart objects"), diseases, environments, festivities, periods of time, and activities with multiple participants are all common categories for PT nouns in the circum-Baltic languages. And we do see PT examples of each of these in Hiaki. Indeed, with a little introspection and attention to literature discussing English (e.g., Wierzbicka 1988; Williams 1994; Acquaviva 2008), we can identify sets of English PT nouns which instantiate similar categories:[9]

(11)
- **i. Substances**:
  Feces, suds, dregs, grits, oats, guts, bowels
- **ii. Multipart objects/complex artifacts:**
  pliers, clippers, binoculars, jeans, khakis, capris, leggings, frills, trimmings, bells and whistles, effects
- **iii. Fabric:**
  clothes, duds, togs, trappings
- **iv. Aggregates:**
  Riches, alms, savings, goods, remains, earnings, spoils, supplies, refreshments, belongings, shavings, smithereens, groceries, oodles
- **v. Words/music:**
  thanks, congratulations, regards, best wishes, compliments
- **vi. Diseases:**
  mumps, measles, shingles, jitters, heebie-jeebies, doldrums, willies, creeps, delirium tremens/DTs
- **vii. Environment:**
  outskirts, surroundings, premises, Rockies, Alps, Himalayas, Great Plains

[9] In fact, as discussed in Acquaviva (2008: 17–18), several nominals in these categories have special lexical content/reference when plural, despite also having a singular form with a more prosaic or literal meaning. Since these do have a singular/plural contrast, they are not strictly speaking 'pluralia tantum', but given the not-strictly-compositional relationship between their singular and plural meanings, it seems likely that these 'lexical plurals' should be counted among our examples in (11) above. Some examples include abstractions over events such as *dreams* ('in your dreams'), *greetings, regrets, looks* ('he has his father's looks'), *funds, wraps, bedsheets*, *(bed)covers*, and *curtains* ('it's curtains for you!').

viii. **Multiples of a situation:**
heroics, shenanigans, high jinks, monkeyshines, antics, airs

ix. **Multipart abstractions:**
odds, auspices, manners, mores, arrears, p's and q's

Considering this data in a cross-linguistic context, we see that what stands out about Hiaki overall is not the particular *categories* of PT nouns, but rather the *size* of the PT inventory. Other language families that have been reported to have a 'large number' of PT are Baltic, Slavic, and Finnic but it is unclear what 'large' means; studies (Koptjevskaya-Tamm & Wälchli 2001; Corbett 2019) look at item lists of a few dozen rather than at the whole nominal inventory as represented in the dictionary. The 22% of Hiaki nouns that are PT seems quite noteworthy in this context.

From a cross-linguistic perspective, it is also worth noting that PT nouns are unusual in the Uto-Aztecan language family (p.c., Friends of Uto-Aztecan Conference 2021, p.c. Kenneth Hill), despite being very robust in Hiaki. This is perhaps surprising given that in languages where PT is well developed, it tends to be diachronically stable (Koptjevskaya-Tamm & Wälchli 2001: 632), and is reconstructed to the proto-language. Although we know that Hiaki PT were documented in the early 1700s, and may thus be diachronically stable, it is surprising that there aren't robust pluralia tantum nominal systems elsewhere in the language family. This aspect of the Hiaki system certainly deserves further investigation within the context of the development of the language family.

3 A closer look at Hiaki PT categories

We now turn to consider the relationship between a nominal's status as a borrowed lexeme and categorization as a PT noun. We argue that it is the semantic category of a borrowed noun that is most relevant to its status as PT, rather than the fact that it is borrowed.

3.1 Generality and productivity of PT categories in Hiaki: Borrowed Ns

The key observation that emerges from our data is that the Spanish borrowings which are marked as PT fall into the same semantic categories as native Hiaki PT. All the PT borrowings in Molina et al. (1999) fall into one of the relevant semantic categories. We illustrate with some salient examples below:

(12) Some PT Spanish borrowings, semantically categorized:

i. **Liquids:**
leechim milk

ii. **Foodstuffs:**
arosim rice; *enchiladam* enchilada; *gayeetom* cookies; *keesum* cheese; *keetim* cake; *luusem* candy; *miisam* host

v. Fabric:
alfonram rug; *avi'itom* habit; *kaapam* cape; *karsetiinim* socks; *kortiinam* curtain; *loonam* canvas tarp; *manteelim* tablecloth

vii. Bichos:
chinchim chiggers, bedbugs, scabies

ix. Hand tools:
paalam shovel; *mache'etam* machete; *hoosom* sickle

x. Multipart objects:
ehkaleam ladder; *ehkalonim* stairs; *kananam* bandolier; *karenam* chain; *kuetem* rockets fireworks; *muevlem* furniture

xiv. Abstract:
koloorim color; *unumpo* at one o'clock

xv. Landscape:
kaayam street

xvi. Groups:
polesiam police

Spanish borrowings that do not denote in a PT-type semantic category are typically not borrowed as PT nouns (13), making it unlikely that grammaticization as a PT noun was a nativization strategy for borrowings.[10] This point is underscored by the fact that most of these words have been fully nativized phonologically and morphologically, and some are not salient as borrowed terms to speakers. For example, *kava'i* 'horse' has undergone considerable phonological change from the Spanish original *caballo*; as have *kus* 'cross' (from Spanish *cruz*), and *kompae* (from Spanish *compadre*).

(13) **A few non-PT Spanish borrowings (there are more)**
anilio ring; *govierno* government; *ehtapia* stamp, *baas* bus; *domisilio* address, *kava'i*, horse; *chiva'a* goat; *ehpeeko* mirror; *kitara* guitar; *kompae* compadre, *kus* cross, *wakas* cow

As noted above, a few borrowings which do clearly denote in one of our PT domains have failed to be grammaticized into a PT category in Hiaki. We list every such case that we have been able to identify with certainty below:

[10] Koptjevskaya-Tamm & Wälchli (2001: 633) in their areal investigation of PT in the Circum-Baltic languages note some cases of borrowing of non-PT nouns into languages with more robust PT categories in which the borrowed noun is treated as PT in the target language. They assert that such cases arise because of the 'formal plurality of the concept', which is 'independent from the etymologic word'. This is consistent with our claim here, namely that the conceptual content of the Hiaki PT category of 'hand tools' is responsible for the application of PT morphosyntax to the Spanish borrowings in this category. A similar point holds for the 'fabric' category, where we see the borrowed PT *kamisetam* from Spanish singular 'camiseta'.

(14) **i.** **Liquid substance**: *kafe* coffee, *kreema* cream
(contrast with borrowed PT *suerom*, IV fluid)
ii. **Foodstuff:**
asuka sugar (contrast with borrowed PT *keesum* cheese)
v. **Fabric:**
korvata necktie (contrast with borrowed PT *paayam* necktie(s))
karpeeta carpet (contrast with borrowed PT *alfonram* rug(s))
ix. **Hand tool:**
kucha'ara spoon
(contrast with borrowed PT *kuchi'im* knife(knives))
xiii. **Words:**
manda vow(s)

In sum, we have seen that the many borrowed words which have been nativized as pluralia tantum nouns fall into the same categories as native PT in Hiaki. We thus claim that their participation in the PT system is largely (or entirely) driven by their semantic content, rather than by their status as borrowings. We observe that one of the reasons that the use of PT marking with borrowed nouns is particularly salient is because two of the relevant semantic domains are areas that saw a lot of borrowing upon contact with colonizing Spanish culture. This includes names for hand tools, where native Hiaki terms like *sena'asom* 'pascola's disk rattle(s)' or *hi'ikiam* 'needle(s)' were presumably already PT before contact, and names for articles of clothing and fabrics, like the native Hiaki terms *hipetam* 'mat(s)' or *sekawam* 'matachin crown ribbon(s)'.

In the following subsection we look more closely at specific categories of PT that might be surprising from an Indo-European perspective: hand tools, constellations and times.

3.2 Unusual pluralia tantum categories: Hand tools, constellations and times

Hand tools, at first glance, may seem to be a surprising PT category: many are prototypically only used one at a time (e.g. *paalam* 'shovel' and *mache'etam* 'machete'). Hand tools are not typical pluralia tantum nouns in Indo-European languages, where most extant studies on PT have focused. The fact that the behavior of these nouns is unexpected from an Indo-European perspective may have even motivated the earlier proposals that PT marking was a borrowing strategy in Hiaki. However, as mentioned above, native Hiaki terms for handheld items are also generally pluralia tantum, even when they are only used in one hand:

(15) a. *ayam* "deer dancer rattle(s)"
b. *hilukiam* "musical rasper(s)"
c. *sena'asom* "pascola's disk rattles"
d. *hi'ikiam* "needle(s)"

Further, the vast majority of Spanish borrowings for hand tools are mapped to the PT category—and as we would expect for this semantic domain, there are many such examples:

(16) A few hand tool PT borrowings
asaroonim hoe(s); *kuchi'im* knife(knives); *martiom* hammer(s); *liimam* file(s); *paalam* shovel; *mache'etam* machete(s); *ehpam* sword(s); *lansam* spear(s), lance(s); *chiivam* crowbar(s); *mule'etam* crutch(es); *reemam* oar(s); *tena'asam* pliers

Why should this semantic domain in particular show such robust PT categorization? We hypothesize that tools that are held in the hand are categorized as PT because *hands* are—*mamam* 'hand(s)', **mama* 'hand'. Whether discussing one hand or many, as for all paired body parts in the language, the plural form is always used. The categorization of paired body parts as PT is cross-linguistically unexceptional. What we propose (based on an idea from Jacqueline Guéron, p.c.) is that tools are, conceptually, *extensions* of the hand. Since hands are canonically PT, so too are tools.[11] This would be an interesting domain for typological exploration—as far as we know, most tools that are marked as PT in Indo-European languages have bipartite internal structure *(scissors, pliers, binoculars*), and it is that internal structure that underlies their membership in the PT category; Hiaki is the first language we have heard of where even non-internally-complex hand tools like shovels are categorized as PT.[12]

Another surprising PT category (from an Indo-European perspective) was the names of constellations, which are PT even when the image and word associated with the constellation is not. For example, one of the names for the constellation referred to in English as 'The Big Dipper'

[11] A reviewer wonders if perhaps the plurality of tool-denoting PT nominals might instead be attributable to their possible status as deverbal nominals and asks whether Hiaki tool words have verbal sources. They compare this conceptually to certain English result nominalizations which seem to be PT nominals, such as *sweepings, belongings, surroundings, scribblings* etc. (see discussion in Mackenzie 2019). In fact, this is not a plausible source for most of these tool nouns; certainly the Spanish-borrowed PT tool nouns like *paalam* 'shovel' or *ehpam* 'sword' do not have a verbal source in Hiaki, and even the native Hiaki PT tool nominals mostly lack any obvious verbal source. The two words for different types of rattle, *ayam* and *sena'asom*, do not correspond to any Hiaki verb we know of, nor does *hilukiam* 'musical rasper'. The only one which does seem to have a native verbal source is the Hiaki word for 'needle': *hi'ikiam*, seems to be derived from the verb *hi'ik* 'to sew'. However, the nominalization process that produced *hi'ikiam* also produces singular count nouns, for example *bwawia* 'sharpened end' is a nominalization of the stative intransitive verb *bwawi* 'be.sharp', but is not a PT noun. Thus, even for *hi'ikiam*, we cannot attribute its PT status to the fact of being deverbally derived via *-a* nominalization.

[12] Wierzbicka (1988: 536) argues of such 'symmetrical action' PT nouns as *scissors, pliers* etc. that their PT status is due to the fact that they are multi-part functional objects, each part of which performs a symmetrical or identical function. Thus, she connects this 'bipartite tool' plurality to the fact that each of the subparts contributes to the function of the tool in the same way. One example that she highlights is Polish PT *skrzypce* 'violin(s)', where each of the four strings of the violin resonate and contribute to the overall function of the violin in the same way. This case illustrates Wierzbicka's point that such PT tool categorization is not dependent on a bipartite structure *per se*, but rather on parallel subparts performing a parallel function. The challenge we see for the Hiaki cases such as *chiivam* 'crowbar' is that there is no obvious internal multipart structure for most of these borrowed tools.

is *kaarom*, a PT word derived from the non-PT Hiaki *kaaro* meaning 'car' or 'cart' (itself a Spanish borrowing, from *carro* 'cart') combined with the plural marker *-m*. The constellation, crucially, is not being described as containing multiple cars or carts; instead, it is the multi-part nature of a constellation—a configuration of multiple stars—that gives it its pluralia tantum status. Similarly, the alternative native Hiaki name for the Big Dipper, *choki araum*, lit. 'star plow-PL', is clearly iconically based on the overall shape of the constellation configuration, rather than on some concept involving multiple plows.

Finally, it is worth mentioning that Hiaki names for specific times in a 12-hour clock are PT as well, and exhibit a few interesting idiosyncracies. Most such time names use native Hiaki numbers, e.g. *woi-m-po* two-PL-at 'at two o'clock', and are unsurprisingly marked plural. However, 'one o'clock' must use the Spanish word for one, (*uno*) as the base, forbidding the use of either of the native Hiaki words for 'one', *senu* or *wepulai*, and furthermore *uno-* must be marked for plural: *(h)uno-m-po* 'at one o'clock'. It's on the basis of this last example that we assume time number names represent a type of abstract PT noun.

4 A Distributed Morphology model of PT in the grammar of Hiaki

In the previous sections we considered how pluralia tantum nouns pattern in Hiaki. Now we turn to the larger implications for theories of pluralia tantum and number marking cross-linguistically. As noted for English by Williams (1994), Hiaki PT nouns give the appearance of being determined by particular lexical semantic properties, but the lexical semantics does not fully and consistently predict pluralia tantum morphology. For example, despite the fact that both *kafe* 'coffee' and *suerom* 'IV fluid' denote liquids, only the latter is plurale tantum. Similarly, despite *veho'orim* 'lizard(s)' and *wikit* 'bird' both denoting small, countable living things, only the former is plurale tantum. This bidirectional irregularity lets us know that we need to be able to model the category of PT nouns arbitrarily as classes in the morphology, rather than grounding PT marking directly in the semantic properties of the nouns themselves. The morphological class of PT nouns is thus like gender (Corbett 2013), in that it seems to be grounded in or developed from semantic categorization but now has an independent life as a grammatical category.[13] Crossing PT status with semantic countability, we can see that all four possible categories of nominal are instantiated in the language (**Table 2**):

We model this behavior within a Distributed Morphology approach to the morphosyntax by appealing to the two different types of number features identified in Weschler & Zlatić (2000; 2003): 'Concord' features, which have morphosyntactic reflexes but are not interpreted, and

[13] A reviewer wonders whether the grammatical class of PT nouns might actually fall under a category of 'gender', or other subclassification of nominals, such as a declension class. Despite the tempting similarity, we do not explore this possibility for Hiaki; see discussion in Corbett (2013: 224–234) for arguments against doing so. In that work, he argues extensively that PT categories are not a subtype of gender or declension class.

	countable (atomic)	**non-countable (mass)**
non-pluralia tantum	*senu wikit* “one bird”	*(*senu) kafe* “(**one*) coffee”
pluralia tantum	*senu tahkai-m* “one tortilla-PL”	*(*senu) vaa’a-m* “(*one) water-PL”

Table 2: Examples of non-PT count, PT count, non-PT mass and PT mass nouns of Hiaki.

‘Index’ features, which are interpreted. We propose that pluralia tantum nouns lexically require the presence of a plural Concord morphosyntactic feature, [–sg], realized in the morphology by the plural *-(i)m* suffix. For PT count nouns, the [–sg] value of the Concord number feature will sometimes contradict a [+sg] value of the Index number feature, which is the interpreted number feature. In this way, semantic number and morphological number can be dissociated from each other. We detail this account in the following sections.

4.1 Number marking in Distributed Morphology

For concreteness, we adopt the Distributed Morphology framework for modeling the morphosyntactic interface. In Distributed Morphology, morphosyntactic features occupy the terminal nodes of a (Minimalist) syntactic tree structure (Halle & Marantz 1993; Siddiqi 2010, a.o.). These features receive phonetic expression via competition by different Vocabulary Items (‘morphs’) when the derivation reaches PF (‘Phonological Form’), and they receive semantic interpretation when the derivation reaches LF (‘Logical Form’). The framework is designed to support full and explicit mappings of morphological representations to semantic representations via the syntax. In the usual case, these representations will stand in a regular one-to-one relationship. For example, in the case of a regular Hiaki count noun, which occupies the bottommost element of a syntactic nominal functional projection containing at least NP and Num,[14] the featural content of the Num head determines both whether that count noun is interpreted as singular (atomic) or plural (non-atomic) on the LF side, and whether that count noun is marked with *-∅* or *-(i)m* on the PF side.

Challenges arise, however, when we consider pluralia tantum count nouns, which we have seen are always marked plural even though they can denote in both the atomic and non-atomic

[14] Nothing that we are saying here hinges on the categorizing role of *n*, attaching to an acategorial root √, so we abbreviate the structure $[[\sqrt{}]n]_{nP}$ as NP for ease of representation. Our abbreviation should not be taken as a claim that the n+√ internal structure is absent, just that it doesn’t matter for the present distinction. There is a locality issue that is worth mentioning — the PT status of the $[[\sqrt{}]n]_{nP}$ depends on the identity of √, which is not local to Num given the intervening n head. Either the relevant [–singular] feature is added to the n head by a rule sensitive to the identity of the √, which is local to √, and is then copied to the local Num head, as in some previous DM analyses, see e.g. Acquaviva (2008); Siddiqi (2009), a.o. or else the PT feature-adding rule applies directly to the Num head, and can be sensitive to √ across the intervening n head. For relevant discussion see Kramer (2016).

domains. We repeat the data from (7) above in (17). Recall that *supem* 'dress' (in 17a) and *tahkaim* 'tortilla' (in 17c) always bear plural inflection, although they can refer to singular or plural dresses/tortillas without issue:

(17) a. *Nee wepul supem hinuk.*
nee wepul supe-m hinu-k
1SG.NOM one dress-PL buy-PFV
"I bought one dress."

b. *Nee wepul hiokareota/*hiokareom hinuk.*
nee wepul hiokareo-ta/*hiokareo-m hinu-k
1SG.NOM one writing.instrument-ACC.SG/*writing.instrument-PL buy-PFV
"I bought one pencil."

c. *Uuchi wepulaim nee ya'aria.*
Uuchi wepulai-m nee ya'a-ria
again one-PL 1SG.ACC make-APPL
"Make me one [tortilla] again." elided PT *tahkaim* "tortilla(s)"

We could model this by hypothesizing that a PT noun like *tahkaim* or *supem* bears an uninterpretable [–singular] feature, which is always realized as *-(i)m* and is accordingly active in the syntax in concord operations, but we then would need to explain why this feature doesn't receive a [–singular] interpretation at LF, the way a regular [–singular] Num head must. In short, why do PT nouns allow this mismatch, where their syntactic number marking and semantic number interpretation are contradictory?

This challenge is compounded by the fact illustrated in (3) above and repeated below as (18) that although syntactically controlled number marking always shows plural agreement with a PT noun regardless of semantic number, number-sensitive suppletive verbs reflect the semantic/interpreted number of their Theme argument,[15] and do *not* reflect its morphological plural number marking.

(18) a. *Hunume livrom ama mesapo vooka.*
Hunu-me livro-m ama mesa-po vo'oka
DET-PL book-PL there table-on lie.SG
"That book is lying on the table."

[15] The argument that controls the form of a suppletive verb can be characterized in a unified way if we state the generalization over theta roles; suppletive verb forms are controlled by the number of an argument that receives a Theme theta role. Subjects of intransitive verbs of body position and motion, and objects of certain transitive verbs, condition suppletion. Agentive subjects of intransitive or transitive verbs, and subjects and objects that are not Themes selected by the suppletive verb itself (e.g. subjects introduced by causatives or objects introduced by applicatives) never condition suppletion. For further discussion and illustration see Harley et. al (2009; 2016) and Bobaljik & Harley (2017).

b. *Hunume livrom ama mesapo to'oka.*
Hunu-me livro-m ama mesa-po to'oka
DET-PL book-PL there table-on lie.PL
"Those books are lying on the table."

That is, interpreted number and actual syntactic number marking must be modeled independently of each other. We turn to our proposal for how to do this in DM in the next section.

4.2 Options for modeling PT number independently of semantic number in DM

To model this chimeric system, we will adopt and adapt an insight from Weschsler and Zlatić (2000)'s treatment of Bosnian/Serbian/Montenegrin/Croatian agreement in Head-driven Phrase Structure Grammar, according to which purely morphological phi-features, called 'Concord' features, are distinct from semantically interpreted phi-features, called 'Index' features. Nominals bear both Index and Concord features, and different classes of agreement phenomena are sensitive to the distinct types of features. This allows an account of 'mixed' agreement patterns within the extended nominal projection of Bosnian/Serbian/Montnegrin/Croatian. (See also Landau 2016 for a Distributed Morphology analysis of a Hebrew PT nominal exploiting Wechsler and Zlatić's feature types).

Plural marking on pluralia tantum nouns, adjectives and determiners in Hiaki is clearly a case of uninterpretable Concord features controlling number agreement throughout the nominal domain. In contrast, we claim that the number-sensitive suppletive verbs are conditioned by the interpretable Index features of their Theme argument. The verbs pay no attention to the Concord features. Unlike the proposal in e.g., Wechsler and Hahm (2011: 259–260) for treating interpretable number marking on the target predicate as semantically active just in case number on the NP [= DP] is underspecified, we locate all semantically interpretable number in the DP itself. In our analysis, predicate number marking is never semantically active, but is conditioned by the syntactic Num features which *are* semantically active, namely Index features. We explicate this distinction between the approaches further below.

We suppose that in the unmarked case, on a regular non-PT count noun, the Concord feature copies the value of the Index feature and thus exhibits uniform behavior across both nominal and verbal number markers—plural-marked nouns co-occur with plural suppletive verbs and vice versa. In the case of countable PT nouns, however, the values for Concord and Index number can come apart. We analyze each of the four logically possible combinations from **Table 2** in turn below, which we summarize again to remind the reader of the categories we identified:

i) 'normal' countable, non-pluralia tantum nouns
ii) countable pluralia tantum nouns
iii) non-countable, non-pluralia tantum mass nouns
iv) non-countable pluralia tantum mass nouns.

To begin, we consider the case of a 'normal' count noun, where morphological number marking in the DP perfectly reflects number interpretation throughout the clause, as in the case of human-denoting nouns like *bwikreo* 'singer'.

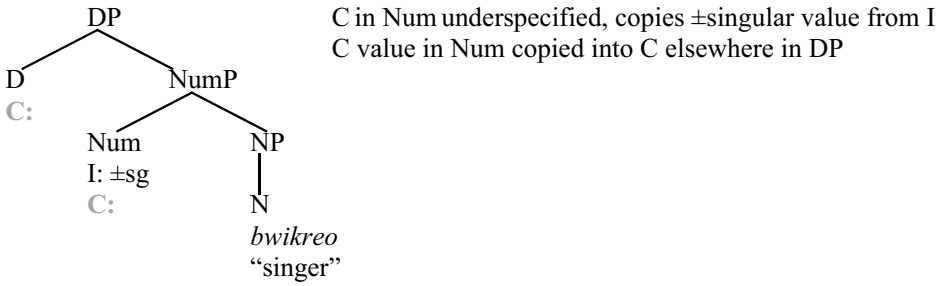
(19) a. *Uu bwikreo aman vuite.*
uu bwikreo aman vuite
DET.NOM.SG singer.SG there run.SG
"The singer is running over there."

b. *Ume bwikreom aman tenne.*
ume bwikreo-m aman tenne
DET.NOM.PL singer-PL there run.PL
"The singers are running over there."

Here, the values of the Concord and Index features in Num line up perfectly.[16] We hypothesize that for such 'normal' count nouns, the Concord features of Num are underspecified and their values are filled in by a feature-copying operation which copies the interpretable Index feature value in Num. This ensures that in the 'normal' case the exponent of Num at PF and the interpretation of Num at LF will match.

The Concord feature in Num then determines the number marking on all elements in the extended nominal projection which bear Concord features. In DM, this is cashed out by a value-copying operation. The Concord feature on e.g. D enters into an Agree relationship with the Concord feature on Num (via Agree Closest, as in e.g. Bošković 2009) and copy its value. (We indicate pre-specified Concord (C:) features in **bold** font, and copied Concord features in **greyscale** font. Index features are given simply as 'I:[±sg]').

(20) A 'normal' count noun:



[16] Note that the verbal exponent is exactly sensitive to the 1 vs >1 distinction, as illustrated by coordinated singular subjects in examples like the following:

i) *Yooko Hoan into Peo tennivae/*vuitivae.*
yooko [Hoan into Peo] tenni-vae /*vuiti-vae
tomorrow [John and Peter] run.PL-PROSP /run.SG-PROSP
'John and Peter are going to run tomorrow.'

We turn now to the mechanics that determine the insertion of the correct verb form. We implement suppletive verb agreement in DM via vocabulary insertion rules of the form in (21) below, as argued in Harley (2014). (See Harley et al. (2016) and Bobaljik & Harley (2017) for detailed discussion and motivation for using the plural form as 'Elsewhere', rather than the singular.) In the case of a 'normal' count noun, of course, we cannot be sure which of the features—Index or Concord—is conditioning the verbal form, since they have identical values, but to anticipate, we include the "I:[+sg]" notation in the conditioning context in the rule, since it will become crucial momentarily:

(21) √RUN ↔ *vuite* / [… NP…]$_{I:[+sg]}$ ——— v
√RUN ↔ *tenne* Elsewhere

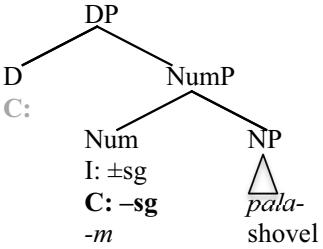
To repeat: when a normal count noun is the Theme argument of a suppletive number-sensitive verb, as in (21), the form that realizes the root matches both sets of features on the noun, so we can't disentangle the question of whether the verb is sensitive to Index or Concord features with this class of nouns.

Let us now turn to the case of countable pluralia tantum nouns like *supem* 'skirt(s)' or *paalam* 'shovel(s)', whose number marking within the DP is entirely and always plural, but for which the verb changes form depending on whether singular reference or plural reference is intended, and where combination with the singular numeral *senu/wepulai* 'one' is grammatical despite the morphological plural marking (as in ex. (7a/17a)).

With such PT nouns, the nominal form requires a plural Concord feature in its Num head. We can implement this within DM with a spell-out rule for the nominal that allows insertion only into structures containing a Num head with value **C: [–sg]**, as illustrated in (22a) below. The Index feature of the Num head, however, is not restricted at all, and can vary between singular and plural according to the speaker's intention, i.e., it is compatible with both I:[+sg] and I:[–sg]. The resulting structure is illustrated in (22b):

(22) a. a countable pluralia tantum noun spell-out rule:
√SHOVEL ↔ *pala* / [$_{NumP}$ [$_{NP}$ __] Num$_{C:–sg}$]

b. countable pluralia tantum noun structure:



C in Num specified as [–sg]
C in Num value copied into C elsewhere in DP
Interpretation at LF conditioned by [±sg] I

(Note that this morphosyntactic structure has the desired effect of representing the *-m* suffix as a normal exponent of a [–sg] Num head, rather than as a substring of the noun itself. This

is necessary to account for the absence of the plural suffix on PT nouns in compounds, as in (6b, d) above.)

To reiterate, while the Index feature on a PT noun varies according to the intention of the speaker, the Concord feature is restricted to only [–sg], and that value is copied to all other locations that probe for Concord number features.

Now we see why it was crucial to include the I:[+sg] specification in the conditioning context for suppletive verb insertion in the rule in (21) above: The verb form of a suppletive verb with a PT Theme noun is conditioned by the value of the Theme's Index feature, rather than by its Concord feature, since the verb form varies according to the intended semantic number.

Now let us turn to uncountable nouns, i.e. mass nouns. We consider first the cases like *kafe* 'coffee' or *ohvo* 'blood' which are marked singular in the nominal domain: *uu kafe* 'the.sg coffee', not **ume kafem*. As expected, such mass nouns condition singular verb forms for suppletive number-marking verbs (like English mass nouns):

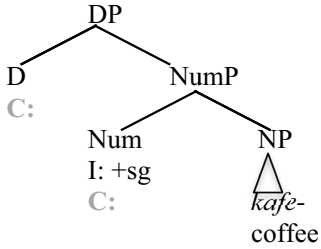
(23) *Uu kafe kom vuitek/*tennek.*
uu kafe kom vuite-k/*tenne-k
DET.NOM coffee.NOM down run.SG-PFV/ run.PL-PFV
"The coffee ran down."

Since such nouns are ungrammatical with plural nominal morphology and take singular verb forms, we can conclude that both their Concord and Index features are marked [+sg]. Indeed, these forms are crucial to our choice to use the Farkas and de Swart (2010) [±sg] feature notation in our analysis of Hiaki number, rather than the [±atomic] number feature often proposed elsewhere in the semantic literature on number (see, e.g. Harbour 2014). These nouns pattern with singulars but cannot be characterized as having a [+atomic] feature, since they denote in the non-atomic domain.

The [+sg] Concord feature on singular mass nouns might arise from the copying of a [+sg] Index feature into an underspecified C: feature slot (as it does for regular count nouns), or it might be required by a nominal vocabulary item rule (as for the plural concord exhibited by countable PT nouns). We will adopt the former idea that it is specified by the copying rule, since the (conceptual) non-pluralizability of mass nouns seems adequate to prevent them from accidentally appearing in the plural,[17] but nothing hinges on this choice.

[17] In preliminary work with our Hiaki consultants, it seems clear that purely inflection-driven 'packaging', as for English *two coffees*, is not available in Hiaki, but more work on the effects of composing plural morphology with singular mass nouns is required before we can say anything definitive. For example, we have not yet been able to establish whether a 'plural kinds' reading is possible with forms like *ume kafem* the.PL coffee-PL, 'the coffees', as it is in English.

(24) singular noncountable (mass) nouns



Since the Index feature conditions verb insertion, the model predicts that the verb form of number-conditioned suppletive verbs with these mass nouns as their Theme will necessarily be singular, as indeed it is.

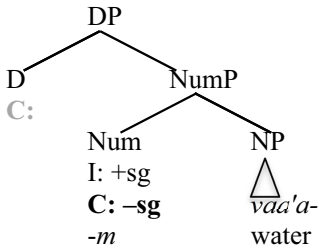
Finally, we turn to the (large) class of noncountable pluralia tantum mass nouns like *vaa'am,* 'water'. Interestingly, like *kafe*, these condition singular verb forms:

(25) *Ume vaa'am kom vuitek/*tenne-k*
ume vaa'a-m kom vuite-k/*tenne-k
the.PL water-PL down run.SG-PFV/run.PL-PFV
"The water ran down."

This behavior entails that they have a [+sg] Index feature in our model. They also must morphosyntactically require a [–sg] Concord feature, as for countable PT nouns, which we implement with the same kind of vocabulary-item spell-out rule, given in (26a). The resulting structure is illustrated in the tree in (26b) below:

(26) a. √WATER ↔ *vaa'a* / [$_{\text{NumP}}$ [$_{\text{NP}}$ __] Num$_{\text{C:–sg}}$]

b. pluralia tantum noncountable (mass) nouns



The observation that the semantically driven agreement indicated by the verb stem is singular even with PT mass nouns confirms our conclusion above: the mass/count distinction in the Hiaki number system cannot be captured using a [±atomic] feature interpreted as such, because the denotation of such non-countable nouns is clearly not [+atomic]. The Hiaki facts require us to create a model where mass nouns and singular count nouns pattern together at the Index feature level, but not at the Concord feature level.

With this class of mass nouns, then, we see a surprising result. Their Concord features pattern with those of plurals, which is consistent with their position at the bottom of the Scale of Individuation, but their Index features, as revealed by suppletive verbal number agreement, are singular. That is, in the features that one would think ought to pattern most closely with their conceptual 'non-individuated' status, they reveal themselves to be singular. It is only in the Concord morphology, which does not bear on their actual interpretive individuation, that we see the patterning suggested by the Scale of Individuation.

It is worth remarking that this behavior is significantly different from that of English, which uses plural marking to indicate number-neutral denotations for count nouns (*Did you buy boxes/*box?*), but in which mass nouns, morphosyntactically similar to plurals in being able to occur without an overt determiner (*I bought a box/boxes/wine*), in fact control singular agreement with demonstratives and the finite verb (*Wine is/*are delicious*; *That/*those wine is delicious*). In applying our model to English, then, mass nouns would seem to be represented with [–singular] Index features but [+singular] Concord features, precisely the opposite of the Hiaki situation.

5 Typological implications

5.1 Hiaki pluralia tantum, verbal suppletion & Corbett's Agreement Hierarchy

In **Table 1** above, we illustrated examples from the extreme ends of Corbett's typological survey of agreement patterns, which led him to propose the Agreement Hierarchy. According to the AH, if a nominal controls agreement on a target higher in the hierarchy (farther to the right of **Table 1**), it will invariably control agreement on targets lower in the hierarchy as well (farther to the left on **Table 1**). So, e.g., plural anaphoric agreement with a given nominal predicts that said nominal will control plural verb agreement and plural nominal concord, since verb agreement and nominal concord are lower on the hierarchy than anaphoric agreement. In contrast, mandatorily plural nominal concord could in principle co-occur with singular verb agreement and singular anaphoric agreement, since they are higher on the hierarchy. Importantly, according to the AH, mandatory verb agreement predicts matching concord agreement, but says nothing about the nature of anaphoric agreement – typologically speaking, nouns exist that look plural, and trigger plural verb agreement, but which allow for singular anaphoric reference when semantically appropriate. Anaphoric agreement is the highest agreement type in the hierarchy.

As noted above, the Hiaki facts pose a challenge to Corbett's Agreement Hierarchy at first glance. Hiaki PT nouns trigger concord on determiners and adjectives (27a), as we have seen, and they are referred to with plural anaphors in subsequent discourse (27b):

(27) a. *Haisa empo ume livrom hinuk?*
Haisa empo **ume** **livro-m** hinu-k
Q 2SG **the.PL** **book-PL** buy-PFV
"Did you buy the book(s)?"

b. *Tuuka ne am hinuk.*
Tuuka = ne **am** = hinu-k
yesterday = 1SG.NOM **3PL.ACC** = buy-PFV
"I bought it/them yesterday."

The Agreement Hierarchy, which places verbal agreement below anaphoric agreement, thus predicts that in Hiaki, verbal agreement should reflect the formally plural status of pluralia tantum nouns.

However, as we have seen, verbs in fact agree according to the actual semantic number of the referent—the value of the Index feature—not according to the formal Concord plural feature (28):

(28)

Escaleam hunum kecha!	*Escaleam hunum ha'abwa!*
escalea-m hunum kecha	escalea-m hunum ha'abwa
ladder-PL there stand.up.**SG**	**ladder-PL** there stand.up.**PL**
"Stand the **ladder** over there!"	"Stand the **ladders** over there!"

This looks prima facie like a problem for the Agreement Hierarchy as stated—we see PT concord agreement and PT anaphoric agreement, but not PT verbal agreement.

We could conclude from this that the hierarchy is different for Hiaki—maybe verbal agreement is exceptionally higher than anaphoric agreement for this language—but in fact, we propose to take a different tack, and argue (with Harley et al. 2016 and Bobaljik & Harley 2017) that Hiaki verbal 'agreement' is not normal agreement, i.e., it is not syntactically mediated via an Agree relation. Only syntactically mediated agreement is subject to the generalizations of the Agreement Hierarchy.[18]

5.1.1 Detour for presentation of Hiaki verbal number

Hiaki verbal number 'agreement' is indicated in only 12–15 verbs of the language and is indexed by choice of suppletive verb form according to the notional singular or plural number of the internal argument, regardless of its grammatical role as subject or object. The pattern of number-conditioned verb suppletion in Hiaki is the typologically familiar one (see Veselinova 2003; 2006): the form of transitive suppletive verbs is conditioned the number of their object, and

[18] It is tempting to think that this is because the Agreement Hierarchy is a side-effect of the syntactic 'distance' between the controller and the agreeing element, given that concord targets are syntactically closer to the controlling head noun than verbal agreement targets, which are, on most analyses, closer to the controlling head noun than anaphoric targets. The intuition that syntactic locality is at play in providing an underlying explanation for the surface patterns of the AH has been expressed often in the literature, see, e.g. Corbett (1979: 216–223). Landau (2016) argues strongly in favor of a syntactic distance account from intervention effects in Index-feature vs Concord-feature controlled adjectival concord in Hebrew. As noted by a reviewer, however, a syntactic distance-based explanation of the whole constellation of AH effects is not obviously plausible given the existence of conflicting agreement behavior from relative pronouns (within a given DP) and matrix predicates (outside it).

intransitive suppletive verbs' form is conditioned by the number of their subject arguments. The intransitive verbs are almost all verbs of bodily position or motion, and hence it is reasonable to hypothesize that their subjects are Themes;[19] the objects of the transitive verbs are also Themes. (See also Harley et al. 2009 for a syntactic argument that the intransitive subjects are not Agents.) A complete list of the suppletive verbs of Hiaki in the speech of our consultants is given in (29)

(29) Hiaki participant-number marking verbs.

Intransitive			Transitive		
Sg. Subj.	Pl. Subj.		Sg. Obj.	Pl. Obj.	
weye	*kaate*	'go, walk'	*kecha*	*ha'abwa*	'stand X up'
vuite	*tenne*	'run'	*yecha*	*hoa*	'set X down'
weama	*rehte*	'walk around'	*kivacha*	*kiima*	'bring X in'
kivake	*kiimu*	'enter'	*me'a*	*sua*	'kill X'
yepsa	*yaha*	'arrive'			
siime	*saka*	'go, leave' (present)			
weche	*watte*	'fall down'			
muuke	*koko*	'die'			
kikte	*hapte*	'stand up'			
yeesa	*hooye*	'sit down' (present)			
vo'ote	*to'ote*	'lying down' (present)			
yehte	*hoote*	'get up'			

Several features of the Hiaki suppletive pattern are worth remarking. First, as illustrated in fn 16, the number distinction is exact: the singular verb form is required with a unit-denoting Theme argument, and the plural verb form is required for a Theme argument denoting any group of two or more entities. We illustrate this again for a transitive verb in (30) below:

(30) a. *Heidi hichikia into palam ama ha'abwak/*kechak.*
Heidi hichikia into palam ama ha'abwa-k/*kecha-k
Heidi broom and shovel there stand.PL-PFV/*stand.SG-PFV
"Heidi stood the broom and the shovel over there."

b. *Heidi uka kovanau kutata ama kechak/*ha'abwak.*
Heidi uka kovanau kuta-ta ama kecha-k/* ha'abwa-k
Heidi the.ACC governor stick–SG.ACC there stand.SG-PFV/*stand.PL-PFV
"Heidi stood the staff over there."

[19] The sole exception, *muuke/koko* 'die.sg/die.pl' also plausibly assigns a Theme theta role to its subject; it's certainly not an Agent, in any case. Suggestively, the plural form *koko* is formally similar to the Hiaki verb meaning 'sleep', *koche* (pl *kokoche*), and the one meaning 'be.sick', *ko'okoe*, perhaps reflecting a diachronic relation to a form connected to meanings involving lying down.

That is, the two forms are in sharp complementary distribution.

Second, the forms are truly suppletive, as diagnosed by a semantic-identity test proposed by Bobaljik (p.c.) and Gribanova (p.c.): identity under ellipsis:

(31) *Identity under gapping: me'a~sua*, 'kill'
Itepo ume toto'im hiva suak, kaa uka kowita.
itepo ume toto'i-m hiva sua-k,
we the.PL chicken-PL just kill.PL-PFV
kaa uka kowi-ta (~~me'a-k~~).
not the.ACC.SG pig-ACC.SG (~~kill.SG-PFV~~)
"We only killed the chickens, not the pig."

(32) *Fragment answers*: *vuite~tenni*, 'arrive'
a. *Havee vuitivae?*
havee vuiti-vae
who run.SG-PROSP
"Who is going to run?"
b. *Jose intok Marcos.*
Jose intok Marcos (~~tenni-vae~~).
Jose and Marcos (~~run.PL-PROSP~~)
"Jose and Marcos." (~~…are going to run.~~)

If the plural forms of these verbs were near-synonyms, rather than truly suppletive realizations of identical underlying semantic content, we would not expect to see the singular form licensing ellipsis of the plural form or vice versa. Compare, for example, the licit elision of *go* under identity with *went* in *John went to the store yesterday but Mary didn't ~~go/*went~~*, and contrast the impossibility of eliding anything but *gather,* yielding an infelicity in verb ellipsis with singular subjects for collective-requiring verbs like *gather* in examples like *The group had planned to gather yesterday but Mary didn't (~~#gather~~)*. Further arguments from idioms, historical patterns, passivization, noun incorporation and speaker-reported metalinguistic observations in support of the notion that these Hiaki number-conditioned verb pairs represent true suppletion were presented in Harley (2015).[20]

Finally, formally singular but notionally plural collective nouns such as *vato'ora* 'people, baptized (ones)' in the Theme position of suppletive verbs require the plural form of the verb, illustrated in (33) below:

[20] This conclusion, that Hiaki verbal number is suppletive and conditioned by 'semantic' number, contradicts Corbett (2000: 258–9)'s assertion that such cases are invariably actually different verbs, comparable to the difference between English *kill* (no number entailments on Theme) vs *massacre* (group number entailments on Theme). See Harley (2015) for the full presentation of this argument.

(33) *Uu vato'ora haivu yahitaite.*
uu vato'ora haivu yahi-taite/*yevih-taite
the.SG.NOM baptized already arrive.PL/*arrive.SG -begin
"The people (lit. 'the baptized (ones)') are already starting to arrive."

This behavior confirms (as does the behavior noted in (28) above with pluralia tantum nouns) that it is 'semantic'/notional number that matters for the conditioning of the correct verb form, rather than formal morphological number.

As proposed in (21) above, we model the choice of suppletive form as morphologically-conditioned allomorphy locally conditioned by the Index features of the sister Theme argument at spell-out, following Harley et al (2016) and Bobaljik & Harley (2017). This type of conditioned morphological spell-out is distinct from 'true' syntactic Agreement in a Minimalist model like DM. True Agreement occurs when phi-feature values (either C-features or I-features, according to the language's grammatical system) on a trigger are syntactically copied and provide values for uninterpretable phi-features on a target. (This is the standard treatment in the literature for concord agreement and for verb-argument agreement, but Agree-based approaches may not be as familiar in the case of syntactic anaphoric agreement. See Paparounas & Akkuş (2023) for an argument in favor of modelling anaphoric agreement via the syntactic Agree relation, as an example of how that could work.)

5.1.2 Hiaki number-sensitive verb suppletion and the Agreement Hierarchy

With this background on Hiaki verbal suppletion in mind, then, let us return to the typological question at hand: Is the behavior of Hiaki suppletive verbs a counterexample to Corbett's Agreement Hierarchy? Recall that we see verb suppletion conditioned by the semantic/notional number of its argument. If this verb suppletion is an instance of proper 'agreement', the Agreement Hierarchy predicts that all types of agreement higher on the hierarchy than verb agreement should also be conditioned by semantic number. In particular, pronominal anaphors should match the semantic/notional number of their antecedent. However, as we have seen, anaphors match the formal number of their antecedent, not the notional number — singular pluralia tantum nouns require plural forms of anaphors, rather than the notional singular forms one would expect. This means that either the Agreement Hierarchy makes the wrong prediction in the case of Hiaki, or the number-conditioned verbal suppletion is not true 'agreement'. In the formal account we have proposed above, we have implemented the latter intuition: Verb suppletion is locally conditioned morphological allomorphy, sensitive to the Index features of the conditioning Theme argument. It is *not* involved in the formal morphosyntactic feature-copying Agree relation. Thus, our analysis not only is consistent with the Agreement Hierarchy, it proposes a specific analytical mechanism for modelling the difference between number-sensitive suppletion and true morphosyntactic agreement.

5.2 Hiaki pluralia tantum and Grimm's Scale of Individuation

We have seen above that the particular nouns which exhibit PT marking fall into recognizable semantic clusters, corresponding to Grimm (2018: 543)'s scale of 'entity types'. As with PT nouns in many languages, it seems intuitively correct to assert that these categories generally have a kind of notional 'natural' plurality to them: PT body parts occur in pairs, constellations are made up of many stars, aggregate substances like rice or beans are made up of many individual grains, etc. These categories, interestingly, correlate with the semantic categories of nouns that exhibit marked singulars—singulatives—in languages that have them, as demonstrated by Grimm (2018).

On the basis of the behavior of marked singular nominals cross-linguistically, Grimm (2018) proposed that nominal referents can be categorized on a 'Scale of Individuation', according the "propensity for the entity described by the noun to occur as an individual", which he connects to the 'accessibility' of the unit interpretation for any given lexical nominal. Marked singulars are nominals that show overt number marking only when they are used with singular reference, rather than showing overt marking when they are used with plural reference, as is the case in languages like, e.g., English. In a typological investigation including Welsh, Maltese, Breton and Dagaare, he showed that marked singulars tend to occur in semantic categories that exhibit conceptual plurality.

Considering the particular semantic content of the nouns which exhibit marked singulars in all these languages, he proposed a conceptual Scale of Individuation, as follows (Grimm 2018: 549, based on a notion in Comrie 1989: 199):

(34) Grimm's Scale of Individuation:
Liquid/Substance < Granular Aggregate < Collective Aggregate < Individual Entities

As usual with such hierarchies, this one is intended to describe the existence of an implicational relationship: Grimm describes this as follows:

> "If two individuation types, say granular aggregates and collective aggregates, stand in a less-than-or-equal-to relationship on the scale of individuation, they will map to categories in the language's grammatical-class inventory that preserve the function."
>
> (Grimm 2018: 550)

This leads us to ask whether the Hiaki PT categories can be insightfully modelled using the Scale of Individuation. Although categories of Hiaki referents that are encoded with PT nominals correspond almost exactly to Grimm's 'entity types' scale, they do not form to a coherent proper subset of individuation types as described by Grimm (2018: 550). The PT nominals of Hiaki span the whole Scale of Individuation, as do the singular nominals. In Hiaki, there exist singular mass nouns, where no unit reference is possible, which are nonetheless morphosyntactically singular,

as in *ohvo* ‘blood’ and *kafe* ‘coffee’. In contrast, we also see countable PT nouns whose referents have a high degree of perceptual individuation, and for which unit reference is available and even likely (see section 6.3 below), like *mache’etam* ‘machete’ or *hi’ikiam* ‘needle’. Indeed, the very countability of PTs—the availability of unit reference as exemplified by quantification by ‘one’—means that (lack of) individuation is precisely not criterial for membership in this ‘plural’ lexical morphological category.

Overall, we conclude that Hiaki pluralia tantum nouns are more usefully characterized by Grimm’s set of entity types than by the Scale of Individuation. Nonetheless, as mentioned above, there is a clear sense in which most of the entity types whose corresponding nominals are in the countable pluralia tantum category do exhibit some kind of ‘notional plurality’ – a group of stars, pairs of body parts, etc. This leads us to ask whether another line of thinking about morphological markedness from the literature could be useful in accounting for the distribution of pluralia tantum nominals in Hiaki. Can we appeal to a frequentist notion of ‘plural reference dominance’ to account for which nouns acquire ‘always plural’ status, morphologically speaking? The Hiaki case presents several features which bear on the debates about predictability and morphological marking in potentially interesting ways.

5.3 ‘Frequentist’ view of morphological markedness in Hiaki PT nouns

The concept of ‘plural reference dominance’ for marked singular categories is construed in terms of predictability of form. On one formulation of this idea, predictability is straightforwardly determined by the frequency of occurrence of a particular form-meaning correspondence in the input (Haspelmath & Karjus 2017). This builds on an interpretation of morphological markedness that is itself dependent on the usual interpretation of Zipf’s Law of Abbreviation (Zipf 1949), encoding the observation that word frequency is inversely correlated with word length: the less frequently a word is used, the ‘bigger’ it is, morphophonologically speaking. Intuitively, if the more frequent form is shorter than the less frequent form, the system minimizes speaker effort while maximizing hearer interpretability: A hearer is much more likely to be able to correctly identify a highly frequent word from a smaller phonological signal than they are to be able to identify a less frequent word from a smaller phonological signal. Therefore, less frequent words should be represented with more phonological material to allow a hearer to unambiguously identify them; conversely, highly frequent words should be represented with less phonological material to minimize speaker effort where it is not needed. Haspelmath and Karjus (2017) argue that ‘uniplex-prominent’ nominals—nouns which are more frequently used in the singular—will therefore show overt coding for plurality, which they call ‘plurative’, while ‘multiplex-prominent’ nominals, which are more frequently used in plural forms, may evolve to show overt coding for singulative.[21]

[21] Haspelmath & Karjus (2017) take the terms ‘uniplex’ and ‘multiplex’ for unit-referring nouns vs collective-referring nouns from Talmy (1988).

This general idea about how morphophonological markedness correlates inversely with frequency has recurred in multiple literatures in various guises. Horn (1984), concerned with scalar implicature, dubbed this tendency the 'division of pragmatic labor': The morphologically marked member of a pair of inflected forms should correspond to the less frequent meaning. Haspelmath et al. (2014) and Haspelmath (2021) dub this the 'form-frequency correspondence hypothesis'.

Mattausch (2007) computationally implements a stochastic OT model which shows that 'bidirectional optimization'—optimization for both hearer and speaker over multiple iterations in a communicative computational model—produces exactly this effect: the morphologically more marked member of a given lexical item with a two-way inflectional contrast is mapped to the less frequent meaning, and the morphologically unmarked member of the pair is mapped to the more frequent meaning.

Haspemath & Karjus (2017) point to the textual frequency of plural forms of translation-equivalent nouns in large corpora in English, Estonian, Latvian, Norwegian and Russian to make the case that singulative forms (in languages that have them) appear with lexical items which tend to be used in the plural more often than in the singular. They find that 18 'singulative-prominent' nominal meanings identified through typological examination of languages with singulative marking are statistically more likely to be used in the plural than in the singular in these large corpora, in contrast with 18 randomly-selected nominals which turn out to be statistically more likely to be used in the singular. Haspelmath & Karjus (2017: 1227) propose a diachronic explanation for the development of singulative forms that is precisely in line with Mattausch's (2007)'s demonstration of a frequency-based evolution of Horn's (1984) 'division of pragmatic labor'. They write "…the correspondence between form and frequency is implemented by diachronic mechanisms which tend to make frequent forms short, because frequent forms are more predictable than rare forms. Ultimately, it is thus predictability that lies at the root of the length difference and the coding asymmetry."

With this background, we can ask whether this line of explanation can offer insight into the development of the coding of particular Hiaki count nouns as pluralia tantum. The pluralia tantum nouns we have documented above tend to belong to semantic categories that are less individuated, yet are morphologically overtly marked as 'plural'. This means that the usual Zipfian idea, where *less*-marked forms should denote plural for nouns with more frequent plural interpretations, does not apply in this case. However, we could imagine a variation on the frequentist idea according to which the *grammaticization* of plural marking on pluralia tantum nouns could be driven by highly frequent plural uses for these nouns. That is, perhaps the grammatical requirement to be always plural developed out of being used more frequently in the plural than in the singular.

This line of thinking as a possible explanation for why the Hiaki pluralia tantum categories are low on the Scale of Individuation runs into trouble, however, in several domains. We cannot

say, for example, that the borrowed Spanish pluralia tantum nouns like *paalam* 'shovel' went through several generations of plural dominant usage and thus finished up as morphologically PT, since they do not have a long history in the language. Or consider *lakim* 'lock', which was likely borrowed directly from English; extensive contact between Hiaki and English did not occur prior to the 1880s, to our knowledge. Another challenge arises in that although mass nouns are extremely low on the scale of individuation, they are not relevantly 'plural', as we saw above from the fact that they condition singular forms of suppletive plural-agreement verbs (section 5.2 above); if anything, we would expect that a frequentist motivation for PT coding would result in singular forms for mass nouns, rather than plural forms.

We can also test the 'plural-reference dominance' approach by adopting Haspelmath & Karjus (2017)'s methodology of quantifying plural-reference dominance for translation equivalents of individual words in a large corpus of another language, making the assumption that these nouns are used with plural reference in that language at reasonably analogous rates in Hiaki.

To arrive at an initial picture of the plausibility of a frequentist view of PT marking on a few individual nouns in Hiaki, we used COCA (the Corpus of Contemporary American English, Davies 2008) to quantify singular and plural frequences for the nouns *wrist, ankle, crowbar* and *machete*, using a proportional sample method.[22] Given a string such as CROWBAR,[23] we would like to know what proportion of the 758 occurrences of that string in COCA are true singular nouns (i.e. eliminating all instances of CROWBAR that are the left-hand member of compounds, present-tense verbs, etc.). We arbitrarily selected 100 matches for CROWBAR in COCA and hand-coded each according to whether they were true singulars or not. In our randomly-selected set[24] of CROWBAR tokens, 92/100 instances were true singular nouns. Assuming that our sample is big enough to be representative of the distribution of singular nominal uses of CROWBAR throughout the corpus, we extrapolate that percentage and arrive at a count of 0.92*758 = ~697 tokens of CROWBAR as a singular noun in COCA. Now, to compare that number to the number of true plural tokens of CROWBARS we repeat the process, examining 100 randomly-selected

[22] It's possible to search by part of speech tags in COCA, which if fully accurate would have given us the ability to report the absolute rather than proportional values of SG:PL ratios for each of our test lemmas, but the part of speech tags seem to be determined in a somewhat shallow way in the corpus, so we found that searches for e.g. ELBOW_nn2 (the tag for N.PL) returned some instances of *elbows* as a verb. Given that we couldn't be sure how or why such erroneous cases are returned, we instead adopted a manually-curated, subsample approach, while recognizing that the resulting values for each category of noun in the corpus (particularly in the singular) are themselves necessarily probabilistic and approximate.

[23] The COCA search we used is not case-sensitive; we represent lemmas in all-caps here to identify the exact letter strings we searched on.

[24] We examined examples #201–300 from the COCA hits for CROWBAR, and individually coded them for true singular count noun use vs 'other' uses. The eight hits which were not singular count nouns included verbs ("crowbar that dough"), exclamations ("Crowbar!") or left-hand members of NN compounds ("the crowbar thing").

instances of CROWBARS to quantify the percentage of tokens in that sample that are true plural nominals. For the string CROWBARS, there were only 86 tokens total, and 86/86 of these tokens were actual plural nominal uses, i.e. 100%. We thus have a singular:plural ratio for the lemma *crowbar* of 697:86—heavily biased in favor of singular uses.

Repeating this procedure for MACHETE yielded 91/100 singular nouns in our randomly chosen sample of 100 tokens of uninflected MACHETE.[25] Thus, multiplying 0.91 times the 1171 total occurences yields ~1066 tokens of MACHETE as a singular count noun. We then searched for MACHETES and examined a randomly chosen sample of 100 hits[26] for that string; 100/100 of these hits were plural nouns, so we estimate that all 498 hits for MACHETES in COCA are true plural uses. Thus, we end up with a sg:pl ratio of 1066:498 for the *machete* lemma, also heavily biased in favor of the singular form.

Taking two other sample words from a different semantic subgroup of our Hiaki pluralia tantum nouns, we compared sg:pl ratios for WRIST and ANKLE using the same methodology. First, for uninflected WRIST, 66/100 tokens in our sample[27] were singular nouns; thus we conclude that approximately ~7731 of the 11,714 tokens of WRIST in COCA are likely to be true singulars. Turning to WRISTS, 100/100 tokens in our sample of WRISTS[28] were true plurals. There were 4585 total occurrences of WRISTS, so we can compute a sg:pl ratio of 7731:4585 for the *wrist* lemma, again showing a notable bias towards the singular use.

Finally, for singular ANKLE, 63/100 instances in our set of 100 tokens were true singulars.[29] Multiplying this percentage by the total number of occurrences of ANKLE yields 0.63*10,394 = ~ 6548 tokens of true singular ANKLE in the corpus. For ANKLES, 100/100 examples[30] were true plural nominals, and there were 4799 tokens total; we thus obtain a sg:pl ratio of 6548:4799, again, showing a singular bias. We summarize these results in the bar chart in **Figure 1**.

For at least these four words, then, it seems quite unlikely that frequency-based plural-reference-dominance could have driven grammaticalization into the PT category.

---

[25] We examined examples #701–800 from the COCA hits for MACHETE. The nine cases that were not singular count noun uses were proper names (“Machete”), left-hand members of NN compounds (“Machete ridge”) or instances of the number-neutral use in the instrumental phrase ‘by machete’.

[26] We examined examples #201–300 from the COCA hits for MACHETES.

[27] We chose to examine examples #5201–5300 from the COCA hits for WRIST. The examples which we did not count as singular count noun tokens included tokens in the collocation “slap on the wrist”, “wrist” as the left-hand noun in NN compounds (“wrist ID band”), and the proper noun “Wrist Deep Productions”.

[28] We chose to examine examples #1001–1100 from the COCA hits for WRISTS.

[29] We chose to examine examples #6901–7000 from the COCA hits for ANKLE. Those matches that were not singular count noun uses were nearly all left-hand members of NN compounds (“ankle bone”); there were no verbal or proper noun uses.

[30] We chose to examine examples #4101–4200 from the COCA hits for ANKLES.

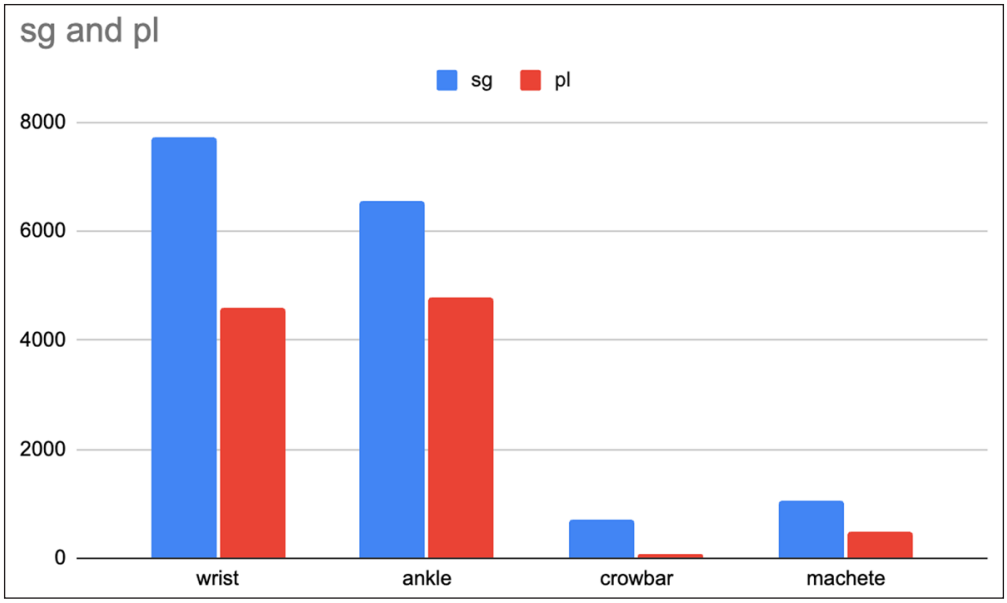


Figure 1: The ratio of approximate singular tokens to plural tokens for four selected English words in COCA whose Hiaki translations are pluralia tantum nouns.

Instead of a frequency-driven process yielding pluralia tantum behavior for individual nominals, then, we conclude that Hiaki PT behavior is better thought of as reflecting membership in a particular semantic class, supporting Kurumada & Grimm 2019's results showing that grammatical coding is more likely to be causally driven by conceptual categorization, rather than by simple token frequency. The particular categories relevant for PT coding in Hiaki, we assume, were already in place as a feature of Hiaki grammar by the time large-scale borrowing of Spanish began, since as we noted above they were documented in the early 17th century Jesuit grammar (Buelna 1890: 44). As Spanish nouns were borrowed, they must then have been slotted into PT or non-PT categories and 'nativized' according to their denotata's membership in the relevant semantic class.

This notion of grammaticization based on semantic category fits with the overall insight of the Concord/Index system for number features proposed by Wechsler & Zlatić (2000; 2003)—the Concord features are 'grammatical' properties, whereas the Index features reflect actual numerosity. As we have seen in the case of pluralia tantum mass nouns, it's the grammatical Concord features, not the semantic Index features, that are involved in marking a mass noun as a plurale tantum noun. The idea that pluralia tantum status is like a noun class feature within Hiaki connected to membership in some language-specific semantic category is consistent with Corbett's notion of noun class features in general being grounded in a nonlinguistic semantic category (gender, shape), which is then extended to large groups of nouns when grammaticized (Corbett 2013).

Finally, when trying to reverse-engineer hypotheses about the sources of PT coding in Hiaki, it is worth reemphasizing that PT behavior is *not* a common property in related Uto-Aztecan languages (p.c. Ken Hill, FOUA 2021 attendees). We assume that this indicates that it is an innovation in Hiaki (and closely-related Mayo), but the path of historical development of this robust pattern seems even more mysterious given that it is unusual for the language family.[31] We hope that future work can shed more light on the diachronic source of this categorial behavior.

6 Conclusions

Overall, 'pluralia tantum' in Hiaki is a much more robust category than suggested in previous descriptions, with hundreds of instances in extant lexicographic documentation. Upon close inspection, the particular nouns which are pluralia tantum fall into a series of recognizable semantic categories, much as pluralia tantum nouns in other languages have been noted to do. We have been able to establish that the borrowed Spanish nouns which are pluralia tantum in Hiaki fall into these same categories as well.

In our formal analysis, we made recourse to the 'Index'/'Concord' number feature distinction proposed by Wechsler & Zlatić (2000; 2003). This enabled us to model the pluralia tantum behavior of mass nouns, despite the fact that they are treated as singular by the verbal suppletive number agreement system. This result has potential implications for the theory of number features, militating against the use of the feature [± atomic], since all mass nouns are semantically singular in Hiaki, although they are not in any way atomic in reference.

The pluralia tantum nouns of Hiaki, since they control both concord and anaphoric agreement but not suppletive verbal number agreement, pose a prima facie challenge to Corbett's typological Agreement Hierarchy, but we showed that in fact suppletive verbal number agreement in Hiaki is not 'true' syntactic agreement (i.e. it doesn't involve functional heads that copy phi-features), and hence the apparent challenge to the Agreement Hierarchy is just that: only apparent.

Finally, the particular semantic categories which pluralia tantum nouns occupy are typically 'low' in the hierarchy established by Grimm's Scale of Individuation. That is, pluralia tantum nouns are more likely to denote in semantic domains with low individuation characteristics, although this is not a requirement. The fact that borrowed Spanish nouns become pluralia tantum according to semantic category suggests that it must be semantic category which motivates grammatical treatment as pluralia tantum, rather than a gradual diachronic grammaticization process.

[31] Recent work (Pasquereau & Henderson 2024) on the Sonoran isolate Seri (ISO 639-3, sei), a close areal neighbor to Hiaki, is showing that all mass nouns in that language are coded for plurality. Perhaps there could have been an influence of Seri on Hiaki favoring marking mass nouns for plurality, which then might have driven grammaticization of plural marking?

Abbreviations

ACC = accusative
APPL = applicative
DET = determiner
GEN = genitive
NOM = nominative
PFV = perfective
PL = plural
Q = question particle
SG = singular
TR = transitive

Data availability

The headword data and semantic organization used in this article is stored with the Open Science Framework and can be found at https://osf.io/uyjm3/?view_only=9380a6f6e8f646f583785f437c86283a.

Ethics and consent

This project was approved by the University of Arizona Institutional Review Board and conducted in accordance with the requirements laid out. The protocol number for this project is 2003435492.

Funding information

This work was conducted with the support of NSF grant BCS 1945641.

Acknowledgements

Many thanks to the other members of the University of Arizona Morphosemantics of Plurals grant team, Robert Henderson, Jérémy Pasquereau (Poitiers), and John Warren William Powell, and audiences at the 2021 Friends of Uto-Aztecan meeting (June 3), the 2022 meeting of the Society for the Study of Indigenous Languages (January 22) and the 2022 Princeton Symposium on Syntactic Theory meeting (April 22). Thanks as well to Guadalupe Vazquez, Lilian Guerrero, Luis Barragan and Michael Everdell for their comments. *Lios enchim hiokoe utte'esia* as well to our Hiaki teachers Maria and Santos Leyva for their years of work helping us to understand their language, and to Homar Aguilar for his work compiling the initial spreadsheet of data on which we base our conclusions here.

Competing interests

The authors have no competing interests to declare.

References

Acquaviva, Paolo. 2008. *Lexical plurals: A morphosemantic approach*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oso/9780199534210.001.0001

Bobaljik, Jonathan David & Harley, Heidi. 2017. Suppletion is local: evidence from Hiaki. In Newell, Heather & Noonan, Máire & Piggott, Glyne & Travis, Lisa (eds.), *The structure of words at the interfaces*, 141–159. Oxford University Press.

Bošković, Željko. 2009. Unifying first and last conjunct agreement. *Natural Language & Linguistic Theory* 27. 455–496. DOI: https://doi.org/10.1007/s11049-009-9072-6

Buelna, Eustaquio. 1890. *Arte de la lengua Cahita, por un padre de la Compania de Jesus*. Imprenta del gobierno federal, Mexico.

Comrie, Bernard. 1989. *Language universals and linguistic typology*, 2nd ed. Oxford: Basil Blackwell.

Corbett, Greville G. 1979. The Agreement Hierarchy. *Journal of Linguistics* 15. 203–224. DOI: https://doi.org/10.1017/S0022226700016352

Corbett, Greville G. 2000. *Number*. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781139164344

Corbett, Greville G. 2003. Agreement: the range of the phenomenon and the principles of the Surrey Database of Agreement. *Transactions of the Philological Society* 101(2). 155–202. Wiley-Blackwell. DOI: https://doi.org/10.1111/1467-968X.00117

Corbett, Greville G. 2006. *Agreement*. Cambridge University Press.

Corbett, Greville G. 2013. Sex-based and non-sex-based gender systems. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *WALS Online* (v2020.3) [Data set]. Zenodo. Accessed at http://wals.info/chapter/31on2024-02-09.

Corbett, Greville G. 2019. Pluralia tantum nouns and the theory of features: a typology of nouns with non-canonical number properties. *Morphology* 29. 51–108. Springer. DOI: https://doi.org/10.1007/s11525-018-9336-0

Corbett, Greville G. 2022. The Agreement Hierarchy revisited: The typology of controllers. *Word Structure* 15(3). 181–225. DOI: https://doi.org/10.3366/word.2022.0208

Corbett, Greville G. 2023. The Agreement Hierarchy and (generalized) semantic agreement. *Glossa: a journal of general linguistics* 8(1). DOI: https://doi.org/10.16995/glossa.9164

Davies, Mark. 2008. *The Corpus of Contemporary American English*. www.english-corpora.org/coca/.

Dedrick, John M. & Casad, Eugene H. 1999. *Sonora Yaqui language structures*. Tucson: University of Arizona Press. DOI: https://doi.org/10.2307/j.ctv1jf2crm

Estrada Fernández, Zarina & Guerrero, Lilian. 2007. Grammatical borrowing in Yaqui. *Empirical Approaches to Language Typology* 38. 419–434. DOI: https://doi.org/10.1515/9783110199192.419

Estrada Fernandez, Zarina & Buitimea Valenzuela, Crescencio & Gurrola Camacho, Adriana Elizabeth & Castillo Celaya, Maria Elena & Carlon Flores, Anabela. 2004. *Diccionario yaqui-espanol y textos: obra de preservación lingüística*. Hermosillo, Sonora, Mexico: Universidad de Sonora; Mexico City: Plaza y Valdés Editores.

Farkas, Donka & de Swart, Henriëtte E. 2010. The semantics and pragmatics of plurals. *Semantics & Pragmatics* 3. 1–54. DOI: https://doi.org/10.3765/sp.3.6

Grimm, Scott. 2018. Grammatical number and the scale of individuation. *Language* 94(3). 527–574. DOI: https://doi.org/10.1353/lan.2018.0035

Harbour, Daniel. 2008. Morphosemantic number: From Kiowa noun classes to UG number features. Dordrecht: Springer. DOI: https://doi.org/10.1007/978-1-4020-5038-1

Harbour, Daniel. 2014. Paucity, abundance, and the theory of number. *Language* 90(1). 185–229. DOI: https://doi.org/10.1353/lan.2014.0003

Harley, Heidi. 2014. On the identity of roots. *Theoretical Linguistics* 40(3–4). 225–276. DOI: https://doi.org/10.1515/tl-2014-0010

Harley, Heidi. 2015. On diagnosing root suppletion. Keynote plenary presentation, *51st annual meeting of the Chicago Linguistics Society*, University of Chicago, April 23, 2015.

Harley, Heidi, & Tubino Blanco, Mercedes & Haugen, Jason. 2009. Applicative constructions and suppletive verbs in Hiaki. *Rice Working Papers in Linguistics* 1. 42–51.

Harley, Heidi, & Tubino Blanco, Mercedes & Haugen, Jason. 2016. Locality conditions on suppletive verbs in Hiaki. In Gribanova, Vera & Shih, Stephanie (eds.), *The morphosyntax-phonology connection: Locality and directionality at the interface*, 91–112. Oxford: OUP. DOI: https://doi.org/10.1093/acprof:oso/9780190210304.003.0004

Haspelmath, Martin & Calude, Andreea & Spagnol, Michael & Narrog, Heiko & Bamyaci, Elif. 2014. Coding causal–noncausal verb alternations: A form–frequency correspondence explanation. *Journal of Linguistics* 50(3). 587–625. DOI: https://doi.org/10.1017/S0022226714000255

Haspelmath, Martin & Karjus, Andres. 2017. Explaining asymmetries in number marking: Singulatives, pluratives, and usage frequency. *Linguistics* 55. 1213–35. DOI: https://doi.org/10.1515/ling-2017-0026

Haspelmath, Martin. 2021. Explaining grammatical coding asymmetries: Form-frequency correspondences and predictability. *Journal of Linguistics* 57(3). 605–633. DOI: https://doi.org/10.1017/S0022226720000535

Haugen, Jason & Harley, Heidi. 2012. Head-marking inflection and the architecture of grammatical theory: Evidence from reduplication and compounding in Hiaki (Yaqui). In Bischoff, Shannon & Cole, Deborah & Fountain, Amy V. & Miyashita, Mizuki (eds.), *The persistence of language: Constructing and confronting the past and the present in the voices of Jane H. Hill*, 133–174. Amsterdam: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/clu.8.06hau

Horn, Laurence. 1984. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In Schiffrin, Deborah (ed.), *Meaning form and use in context: Linguistic applications*, 11–42. Washington DC: Georgetown University Press.

Koptjevskaya-Tamm, Maria & Wälchli, Bernhard. 2001. The Circum-Baltic languages. An areal-typological approach. In Dahl, Östen & Koptjevskaja-Tamm, Maria (eds.), *Circum-Baltic Languages: Volume 2: Grammar and Typology*, 615–750. John Benjamins Publishing Company. DOI: https://doi.org/10.1075/slcs.55.15kop

Kramer, Ruth. 2016. A split analysis of plurality: Number in Amharic. *Linguistic Inquiry* 47. 527–559. DOI: https://doi.org/10.1162/LING_a_00220

Landau, Idan. 2016. DP-internal semantic agreement: A configurational analysis. *Natural Language and Linguistic Theory* 34. 975–1020. DOI: https://doi.org/10.1007/s11049-015-9319-3

Mackenzie, J. Lachlan. 2019. Is there a pluralia tantum subcategory of nominal gerunds? Developing Gaeta's notion of morphological differentiation. *Language Sciences* 73. 179–189. DOI: https://doi.org/10.1016/j.langsci.2018.08.015

Mattausch, Jason. 2007. Optimality, bidirectionality & the evolution of binding phenomena. *Research on Language and Computation* 5. 103–131. DOI: https://doi.org/10.1007/s11168-006-9018-7

Molina, Felipe S. & Shaul, David Leedom & Valenzuela, Herminia. 1999. *Yoeme-English/English-Yoeme: With a comprehensive grammar of Yoeme language*. New York: Hippocrene Books.

Paparounas, Lefteris & Akkuş, Faruk. 2023. Anaphora and agreement in the Turkish DP. *Natural Language & Linguistic Theory*. DOI: https://doi.org/10.1007/s11049-023-09583-4

Pasquereau, Jérémy & Henderson, Robert. 2024. Pluralia tantum nouns in Seri. Paper presented at *An areal exploration of plural systems of the indigenous languages of Northwest Mexico and the Southwest United States* workshop, January 4, 2024, SSILA 2024, New York, US.

Siddiqi, Daniel. 2009. *Syntax within the word: Economy, allomorphy and argument selection in Distributed Morphology*. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/la.138

Siddiqi, Daniel. 2010. Distributed Morphology. *Language and Linguistics Compass* 4. 524–542. DOI: https://doi.org/10.1111/j.1749-818X.2010.00212.x

Veselinova, Ljuba. 2003. *Suppletion in verb paradigms: Bits and pieces of the puzzle*. PhD. Stockholm, Sweden: Stockholm University.

Veselinova, Ljuba. 2006. *Suppletion in verb paradigms: Bits and pieces of the puzzle*. Amsterdam: John Benjamins Publishing Co. DOI: https://doi.org/10.1075/tsl.67

Wechsler, Stephen & Hahm, Hyun-Jong. 2011. Polite plurals and adjective agreement. *Morphology* 21. 247–281. DOI: https://doi.org/10.1007/s11525-010-9168-z

Wechsler, Stephen & Zlatić, Larissa. 2000. A theory of agreement and its application to Serbo-Croatian. *Language* 76. 799–832. DOI: https://doi.org/10.2307/417200

Wechsler, Stephen & Zlatić, Larissa. 2003. *The many faces of agreement*. Stanford: CSLI Publications.

Wierzbicka, Anna. 1988. *The semantics of grammar*. John Benjamins Publishing. DOI: https://doi.org/10.1075/slcs.18

Williams, Edwin. 1994. Remarks on lexical knowledge. *Lingua* 92. 7–34. DOI: https://doi.org/10.1016/0024-3841(94)90336-0

Zipf, Gregory K. 1949. *Human behavior and the principle of least effort*. Addison-Wesley Press.