

# CORRELATES TO MIDDLE MARKING IN DENA'INA ITERATIVE VERBS<sup>1</sup>

Andrea L. Berez and Stefan Th. Gries

University of California, Santa Barbara

## **Abstract**

While recent studies of Athabaskan middle verbs have attempted to find a unified semantic motivation for their presence, middle marking in iterative verbs, which are a subset of middle verbs, is generally less well understood than in other middle constructions. Scholars have cited syntactic intransitivity, semantics, or some combination thereof as motivation for when iteratives are marked as middles. In this paper, we present a quantitative analysis of iteratives from traditional Dena'ina (Athabaskan, Alaska) narratives. This analysis strongly suggests that, while grammatical transitivity plays a role in the triggering of overt morphological marking of middles, the verb meaning plays an even more important overall role and, thus, supports the assumption of a semantically unified class of middle verbs. More specifically, we show that in Dena'ina iterative verbs, middle marking is more likely to occur when the spatial starting and ending points of the action of the verb are undifferentiated.

## **Key words**

Dena'ina, iterative verbs, middle voice, transitivity.

## 1. Introduction

Recent studies of Athabaskan middle verbs have attempted to find a unified semantic motivation for their coherence as a class (Kibrik 1996; Thompson 1996; Rice 2000 *a*; Holton 2000; 2005). In Athabaskan languages, iterative verbs are sometimes middles, but as a set they are generally less well understood than other middle constructions. Scholars have cited syntactic intransitivity, semantics, or some combination thereof as motivation for when iteratives are marked as middles (cf. below for sources). In this paper, we present a quantitative analysis of iteratives from traditional Dena'ina (Athabaskan, Alaska) narratives that strongly suggests that while both intransitivity and semantic distinctions play a role in the triggering of overt morphological marking of iteratives as middles, semantic distinctions are more highly correlated with middle marking. More specifically, we show that in Dena'ina iterative verbs, middle marking is statistically most likely to occur in iteratives that are *spatial reversives*, in which the locational source and goal of the action are undifferentiated. We believe this finding is congruous with Rice's observation that middles in Athabaskan cohere as a class based on the "suppression of differentiation of arguments and suppression of differentiation of events" (2000 *a*: 179). Our analysis thus supports the assumption of a semantically unified class of middle verbs.

The remainder of this introduction provides a brief description of the relevant aspects of Dena'ina morphology, summarizes the previous research into iterative verbs and middle constructions in Athabaskan, and introduces our research questions. Section 2 is an account of the methodology we used in our quantitative study, the results of which are presented in Section 3. Section 4 discusses the implications of our results for unifying Dena'ina iteratives with other Athabaskan middles constructions and concludes.

In the present study, we are considering verbs that contain the iterative prefix *nu-*, as in

*ndunuidyu* 's/he came in again'.<sup>2</sup> As we will see in detail, this morpheme has meanings beyond the multiple iteration of an event. It can also refer to a single repetition of an event or a return to a location or state, and it is sometimes found in verbs denoting an action that takes place on a customary basis.

The term *middle* is used in Athabaskanist literature as an umbrella category for a range of constructions including reflexives, reciprocals, self-benefactives, verbs with body-part incorporates, and the like (cf. Rice 2000 *a* for a summary). All of these constructions contain the marker of middle voice, a morpheme known in Athabaskan linguistics as the *d*-classifier, also referred to here as simply *d*.<sup>3</sup> The *d*-classifier is one of "a four-element set of morphemes, one of which necessarily appears in every occurrence of every verb ... As is generally recognized, the function of the classifiers, if any, is related to transitivity marking" (Kibrik 1996:259). In other words, every verb theme has a lexically specified classifier (in Dena'ina, one of *d*, *l*, *l*, and  $\emptyset$ ), and in many cases speakers can choose a different classifier to indicate changes in voice, valence and/or transitivity. The vast literature on the historical development and behavior of the classifiers will not be reproduced here (see for example Goddard 1905; Hoijer 1946; Krauss 1969; Howren 1971; Kari 1979; Leer 1979; Collins 1979; Thompson 1989; 1996; McDonough 1989; Kibrik 1993; 1996; 2002; Arce-Arenales, Axelrod and Fox 1994; Rice 2000 *b*); for the present discussion it is relevant to know that *d* is generally regarded as indicating a decrease in the semantic transitivity (in the sense of Hopper and Thompson 1980) of a proposition, and that phonology sometimes obscures the presence of the morpheme (although in many cases, segmental features indicating the underlying presence of *d* are identifiable; see Section 2.2.1 below). For this paper, we are adopting Rice's (2000 *a*) sense of the term *middle*, who, while acknowledging that verb forms with *d* do not necessarily reflect the traditional notion of changes

to argument structure, still refers to them as middles.<sup>4</sup> This is essentially a morphological definition of the category, in that inclusion in the class requires the presence of the *d*-classifier.

Scholars (e.g., Kibrik 1996; 2002; Thompson 1996; Rice 2000 *a*) have sought to understand the motivation for marking constructions like reciprocals, reflexives, and body-part incorporates as middles. Rice, building on Kemmer (1993), looks to the semantics of the situations and events described by these constructions and offers the low differentiation of arguments and events as a motivation:

[Kemmer] defines the property of relative elaboration of events as 'the degree to which the facets in a particular situation, i.e. participants and conceivable component subevents in the situation, are distinguished' (1993:208). Kemmer further notes that the 'speaker has a choice of either marking reference to events as undifferentiated wholes, or making reference to their substructure or component parts' (208) ... [T]wo-participant events include an initiator and an endpoint which are distinct from one another while single-participant events involve one participant. Middles typically involve two participants which are not differentiated ... in the middle voice, one finds both suppression of differentiation of arguments and suppression of differentiated events. (Rice 2000 *a*:179)

In most middle constructions, the presence of *d* is predictable and conforms to Rice's account of "suppression of differentiation of arguments," as illustrated in examples (1) to (5). In reflexives and self-benefactives, the participant is both the initiator and the recipient of the action; in reciprocals and indirect reciprocals, multiple participants are initiators and recipients;

in verbs with incorporated body parts, the recipient is a subpart of the initiator. In (1) to (5), the (a) examples are the non-middle forms, and the (b) examples are the middle versions; examples from other Northern Athabaskan languages are given in cases where clearly illustrative examples from Dena'ina were not available to us.

(1) Reflexive: Slave (Rice 1989; in Rice 2000 *a*:180)

- a. *dahyed̥ɪlu* 'she hung it'
- b. *dah'eded̥ɪdlu* 's/he hung him/herself'

(2) Self-benefactive: Dena'ina (Tenenbaum 1978: 126)

- a. *shu dak'inil* 'pour (e.g., coffee) for me'
- b. *dak'did̥nil* 'pour (e.g., coffee) for yourself'

(3) Reciprocal: Koyukon (Thompson 1996:355; in Rice 2000 *a*:181)

- a. *yetots'eyh* 's/he will pinch him/her once'
- b. *neelheetodets'eyh* 'they will pinch each other once'

(4) Indirect reciprocal: Dena'ina (Tenenbaum 1978: 126)

- a. *vet'uy dayeshyu* 'I started walking toward him/her'
- b. *nilt'uy hdasdyu* 'they started walking toward each other'

(5) Incorporated body part: Tsuut'ina (Cook 1984:136; in Rice 2000 *a*:184)

a. *sítsì dìnìsìs'ó* 'I turned my head' (no incorporate)

b. *digá tsìdìnìsít'ó* 'I turned my head' (incorporate)

Turning to iterative constructions in Dena'ina, we find that unlike the constructions above, they are not invariably marked as middles. It has been claimed in research across the Athabaskan family that the occurrence of the *d*-classifier in verbs containing the iterative morpheme *nu* (or cognate) patterns according to transitivity: *d* is said to nearly always occur in intransitive iterative verbs, and is claimed to occur only very rarely, if at all, in transitives (cf. Tenenbaum 1978 for Dena'ina; Rice 1989 for Slave; Kari 1990 for Ahtna; Thompson 1996 for all branches of Na-Dene; Jetté and Jones 2000 for Koyukon; Rice 2000 *a* for general Athabaskan). Rice (2000 *a*:188) summarizes this pattern in her discussion of voice and valence in the Athabaskan family: "[w]hen [the iterative marker is] combined with intransitives, middle-voice marking is found; it is unusual to find this with transitives." The Venn diagram in Figure 1 illustrates the partial inclusion of iteratives (all of which have *nu-*) in the class of middles (i.e., only some iteratives have *d*).

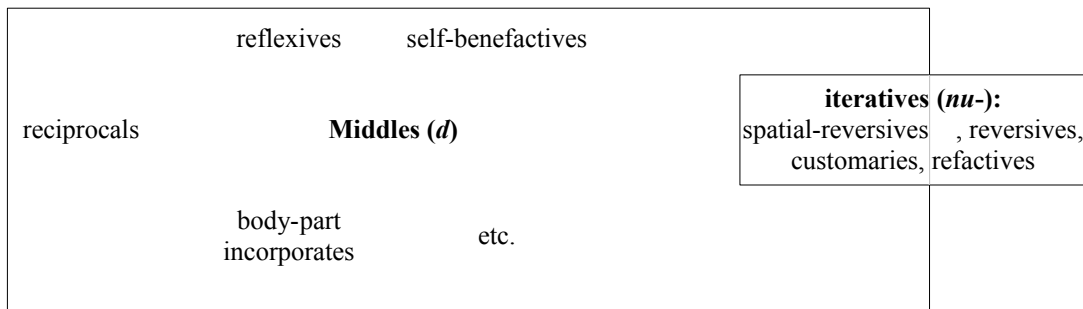


Figure 1: Inclusion of various constructions in the formal class of middles

As we will see, the transitivity-based pattern of the distribution of *d* discussed in the literature holds true for many of iterative constructions in our data. Consider the Dena'ina examples in (6) and (7), both from Tenenbaum (1978: 172). Example (6a) is an intransitive verb with a corresponding iterative construction in (6b), in which the *d*-classifier appears. (7a) is a transitive verb, and *d* does not appear its iterative counterpart in (7b).

(6) a. intransitive non-iterative:

*ghe- yul*

CONJ-walk

's/he is walking'

b. intransitive iterative:

*nu- ghe- d yul*

ITER- CONJ-D.CLF-walk

's/he is walking back'

(7) a. transitive non-iterative:

*yi- t- jeh*

OBJ-CLF-hit

's/he hit him/her (once)'

b. transitive iterative:

*nu- yi- t- jeh*

ITER- OBJ-CLF-hit

's/he hit him/her again'

However, closer inspection shows that in a substantial number of *nu*-forms in Dena'ina, the picture is not as clear as previously thought. There are numerous occurrences of intransitive *nu*-verbs without the *d*-classifier, and there are also transitive examples that do contain it. Compare (6) and (7) above with (8) and (9).

(8) intransitive iterative, without *d*:

*nu- ghi- t- ghatl'*

ITER-CONJ-CLF-become.dark

'it got dark again' (TDS *b*:34)

(9) transitive iterative, with *d*:

*qenaga qa- na- t- de- z- ghe- sh- d- nik*

voice ADV-ITER-INCORP-GEN-CONJ-ASP-1SG.SUB-D.CLF-hear

'voice I'll never hear again' (TDS *d*: 23)

(cf. *qadandeshnex* 'I hear you' (without *d*) (Kari n.d.))

This is not to say that there has been no discussion in the literature of either the need for more nuance when describing the distribution of middle marking in iteratives, or of the potential explanatory power of semantics in understanding when iteratives receive *d*. Holton (2000:189) writes of Tanacross iteratives, "the D- morpheme may occur with both transitive and intransitive

verbs. This is in contrast to some other languages which require the D- only with intransitive verbs." He offers a suggestion of agent-affectedness as a semantic motivation for the absence of *d* in (10) and its presence in (11) below (note the portmanteau of subject and classifier in (11)) such that the agent is presumed to be more affected in (11) than in (10) because in (11) the object returns to the source with the agent. Both examples are transitive verbs from Tanacross (Holton 2000:190).

(10) *ežeg nînihʔa*

ežeg ní- na- n- ih- ʔaʔ

there TERM-ITER-M-1SG-handle.cpt.object:PERF

'I brought it back there'

(11) *jah nanišqʔ*

jah na- n- iš- ʔaʔ

here ITER-M-1SG:D-handle.cpt.object:PERF

'I brought it back here.'

Rice (2000 *a*:188) comments on a semantic commonality between intransitive iteratives on the one hand, and reflexives and reciprocals on the other: "[t]he iterative construction with intransitives is parallel to reflexives and reciprocals, indicating a common source and goal." Thompson (1996:364) also suggests a semantic factor, observing that at least for iteration in a spatial dimension, the source and goal are the same. However, he still clearly maintains that *d* occurs in iteratives in all three branches of Na-Dene "only when the verb is intransitive."

Kibrik has perhaps been strongest in asserting the semantically-based nature of transitivity in Athabaskan. He writes of the function of the classifiers, "all processes invoking a classifier shift can be subsumed under the category of semantic transitivity, as defined by Hopper and Thompson 1980" (1996:260). In the same article he goes on to describe passives and semipassives, in which the semantic parameter of actor-affectedness is accompanied by a morphological indication of transitivity decrease. Kibrik (2002: 8) further notes that detransitivization in oblique reflexivization, "indicates that Athabaskan transitivity is of the semantic cluster type, in the spirit of Hopper and Thompson (1980), rather than of a strictly syntactic character."

Thus the stage is set for a semantic explanation of which iteratives are marked as middles, and which are not. With this in mind, however, we still do not yet have a clear picture – nor even a simple text count – of the distribution of *d* in iteratives for any single Athabaskan language, and until now we have been relying on impressionistic descriptors like *rarely*, *often*, and *frequently*. While impressionistic frequencies can be the foundation on which research programs are built, in this case we believe that a quantitative investigation of the morphological behavior of iteratives can shed new light on the thorny question of what can be regarded a consistent predictor of middle status. This paper investigates the hypothesis that given that a satisfying semantic motivation – Rice's low differentiation of arguments and events – can be found for most middle constructions, middle status in iteratives can be semantically motivated. Finding a sense-based justification for why some but not all iteratives are marked with the *d*-classifier would allow for a unified motivation of the class of middles as a whole. It is this sense-based justification that we investigate here. More specifically, we have two interrelated goals:

- a quantitative-descriptive goal: what is the rate of occurrence of the middle marker *d* in iterative verbs in Dena'ina? What is the proportion of iteratives that are also middles to iteratives that are not middles, and how are these distributed in terms of both transitivity and semantics?
- an exploratory goal: does either of the two kinds of determinants discussed so far – semantics or syntactic transitivity – play a more important role for the presence or absence of *d*? Furthermore, can we discover effects at a finer level of resolution (i.e., effects for different verb senses)?

The semantic nature of transitivity in Athabaskan implies that the presence or absence of *d* is influenced by more than one variable, so the questions we examine here are inherently multifactorial. Because intuition alone is often a poor guide to answering multifactorial questions, we have conducted a corpus-based quantitative analysis of the distribution of *nu* and *d* in Dena'ina texts, the implementation of which is discussed in the following section.

## **2. Methods**

### *2.1 The corpus and the data extraction*

To meet our quantitative-descriptive goal regarding the rate of occurrence of *d* in Dena'ina iteratives, the first author compiled 500 tokens containing *nu* from Tenenbaum (1976 *a-d*), a collection of 24 traditional stories in the Inland dialect of Dena'ina with English word-level glosses and English free translations. The collection of forms was exhaustive rather than selective, and spanned approximately 2,800 lines of text over 220 pages. Our analysis is based on

token frequency.

## 2.2 Annotation

The variables that were included in the analysis were

- MORPHEME: *+d* vs. *-d*.
- TRANSITIVITY: *intransitive* vs. *transitive*;
- SENSE: *refractive* vs. *reversible* vs. *spatial reversible* vs. *customary*<sup>5</sup>

Below we characterize the variable levels and the way they were coded in more detail.

### 2.2.1 The variable MORPHEME

As to the variable MORPHEME, we needed to determine whether or not *d* was present in each token. As mentioned above, phonology can obscure the presence of a derived *d*, and we cannot rely on surface form alone to tell us when one is present. In Dena'ina, *d* coalesces with the underlying lexically-determined classifier, such that the voicing and obstruence features of /d/ surface in the classifier position; additionally, the classifier may not surface as /d/ because of other phonological processes that simplify consonant clusters. The resulting forms vary depending on the original morpheme in the classifier position, as shown in Table 1.

Coalescence of <i>d</i> with lexical classifier ...		... resulting form:
∅	→	[d]
ɬ	→	[l]
l	→	[l]
d	→	[d]

Table 1: Effects of *d* on classifier surface forms

Themes with a lexical ∅- or *l*-classifier show a detectable change when in the presence of *d*, and these forms are realized as [d] and [l] respectively. In themes with a lexical *l* or *d*, however, it is impossible to determine whether or not the *d* indicating middles (i.e., a *productive d*) is present. Once we found the lexically specified classifier for each theme in Kari (n.d.), we discarded 220 tokens with a lexical *l*- or *d*-classifier from the study. This reduced the set to 280 verb forms, namely 227 with productive *d* (+*d*) and 53 forms without *d* (-*d*).<sup>6</sup>

### 2.2.2 The variable TRANSITIVITY

As to the variable TRANSITIVITY, we used Kari (n.d.) and the Tenenbaum texts as a guide for coding: themes with an obligatory object position, or those with an overt NP object were coded as transitive. All others were considered intransitive. Of the 280 forms, 232 were intransitive and 48 were transitive.

### 2.2.3 The variable SENSE

As to the final variable, SENSE, we distinguished four different senses based on semantic subclasses of iterativity (e.g., Cusic 1981; Xrakovskij 1997; Wood 2007; Shluinsky 2007) that were present in the Dena'ina data. Because we are looking at behavioral divisions within the

class of iteratives, we chose to annotate the data according to semantic subcategories specific to iterativity, rather than to according to some other semantic division. Below, we briefly define each sense and provide examples from the forms included in the study.

- (12) Refactives: these forms indicate a second or subsequent occurrence of an event.

*tleg* **nuhtazdlach** 'they started cooking fat again' (TDS *b*: 7)

*qenaga qanatdezgheshd***nik** 'voice I'll never hear again' (TDS *d*: 23)

**nutashchagh** 'he started to cry again' (TDS *a*: 59)

*k'kalt'a* **nuk'ehnilchut** 'they cut off one fish tail again' (TDS *b*: 3)

*des***nuk'ehghildatl'** 'they put more wood on the fire' (TDS *b*: 4)

**nughilghatl'** 'it got dark (again)' (TDS *b*: 24)

- (13) Customaries: these forms refer to events repeated on a customary or regular basis.

**ch'anadyux** 'he customarily goes out' (TDS *b*: 3)

*shtunughed***nex** 'he goes out hunting by boat all the time' (TDS *b*: 18)

**nuqeyltih** 'they used to wash him' (TDS *d*: 17)

**nuhk'ghelax** 'they would make him a swing (regularly)' (TDS *c*: 51)

- (14) Reversives: these forms indicate a return to a prior state.

*qut'an* **nusdlan** 'he became a person again' (TDS *b*: 42)

*yagheli* **nugheshd***lahi* 'I'll be well (again)' (TDS *c*: 79)

**nu***shich'hulchin* 'he got his wind back (i.e., recovered from exertion)' (TDS *d*: 22)

**nu***hustl'in* 'he got dressed' (TDS *c*: 58)

*k'nuqadghildatl'* 'he put on his snowshoes' (TDS *a*: 40)

- (15) Spatial reversives: these forms indicate movement in physical space to a previously occupied location.

*nu'idyu* 'he is coming back' (TDS *c*: 2)

*nughudnex* 'he paddled back' (TDS *b*: 19)

*nuhyilchesh* 'they brought it back' (TDS *a*: 7)

*nuyulghel* 'he put him back' (TDS *a*: 78)

*tunushiziltax* 'bring me back up!' (TDS *a*: 3)

*tunuqesdatl'* 'they came back up' (TDS *b*: 32)

*nu'ihdal* 'you guys come back again' (TDS *d*: 3)

*qananlggat* 'move the boat in! (i.e., back in to the shore)' (TDS *d*: 40)

The 280 tokens included in the study are distributed as represented in Table 2.<sup>7</sup>

	intransitive		transitive		Totals
	+ <i>d</i>	- <i>d</i>	+ <i>d</i>	- <i>d</i>	
<b>refactive</b>	38	19	5	10	72
<b>customary</b>	4	0	0	1	5
<b>reversive</b>	14	1	4	8	27
<b>spatial reversive</b>	145	11	17	3	176
<b>Totals</b>	201	31	26	22	280

Table 2: The corpus data

The data is available in an electronic appendix found at [http://URL\\_HERE](http://URL_HERE). These 280

cases are all the cases that were unambiguously codable for all three variables and are analyzed statistically below.

### 2.3 *Statistical analysis*

Given the exploratory goal outlined in Section 1, we decided to use two different statistical methods in our analysis. On a more coarse-grained level, we wanted to find out if either independent variable – TRANSITIVITY or SENSE – has a stronger influence than the other on middle-voice marking, we modelled the relationship between MORPHEME on the one hand and TRANSITIVITY and SENSE on the other hand using a binary logistic regression model with subsequent model selection. The objective here was to see (i) whether both independent variables and their interaction need to be retained in the model and (ii) which degree of predictive accuracy the minimal adequate model could attain.

On a more fine-grained level, we wanted to explore which senses, if any, correlate with middle voice marking and used a statistical approach more tailored to a fine-grained contrastive analysis of the effects of individual senses, a hierarchical configural frequency analysis (HCFA; cf. von Eye 1990). A HCFA is a method for the analysis of multidimensional frequency tables that is conceptually similar to chi-square tests, but it has two main characteristics that set it apart from these. First, a HCFA generates all possible (or all user-defined) sub-tables for the data set in question and tests all of these for significant deviations from expected frequencies. Second, a HCFA not only tests complete tables for significance, but also tests each individual cell – or *configuration* – in each table for significance; cells whose observed frequencies are higher than their expected frequencies are referred to as *types* while cells whose observed frequencies are lower than their expected frequencies are referred to as *antitypes*.<sup>8</sup> Since we are interested in the

morphological marking, we only discuss the sub-tables that contain the variable MORPHEME.

### 3. Results

#### 3.1 Coarse-grained exploration: binary logistic regression

To predict the use of *d*, we began by fitting a model with MORPHEME as the dependent variable and TRANSITIVITY, SENSE, and their interaction (TRANSITIVITY × SENSE) as independent variables. A comparison of this model to one without the interaction showed that the interaction was only marginally significant ( $\chi^2=6.9$ ,  $df=3$ ;  $p=0.075$ ), so following Occam's razor we removed it from the model. Further simplification, however, was not possible: removing either TRANSITIVITY or SENSE resulted in significantly worse performance (removing TRANSITIVITY:  $\chi^2=16.55$ ,  $df=1$ ;  $p<0.001$ ; removing SENSE:  $\chi^2=31.02$ ,  $df=3$ ;  $p<0.001$ ). Thus, the minimal adequate model included the main effects of TRANSITIVITY and SENSE, but no interaction (however, cf. Section 3.2 below). This model provided an intermediate fit to the data (m.l.  $\chi^2=54.07$ ;  $df=4$ ;  $p<0.001$ ; Nagelkerke  $R^2=0.283$ ;  $c=0.787$ ).

A comment regarding the importance of TRANSITIVITY and SENSE is in order. The above chi-square values already suggested SENSE is more important than TRANSITIVITY, and this informal assessment was confirmed when we compared the predictive powers of both independent variables in isolation. TRANSITIVITY alone did not only perform much worse than SENSE alone ( $\chi^2=14.74$ ,  $df=2$ ;  $p<0.001$ ), but it also resulted in a very small overall correlation: for the relationship between TRANSITIVITY and MORPHEME, Nagelkerke  $R^2=0.127$  and  $c=0.65$  while for the relationship between SENSE and MORPHEME, Nagelkerke  $R^2=0.202$  and  $c=0.734$ . This provides strong evidence for the fact that, while TRANSITIVITY does make a

significant overall contribution, SENSE is responsible for a larger share of the predictive power.

Let us now look at the results in more detail, answering the question of how the individual levels of TRANSITIVITY and SENSE affect the presence of  $d$ . We do so by looking at how the levels of TRANSITIVITY and SENSE influence the odds of MORPHEME:  $+d$ . The odds of a particular situation  $s$  are computed as the quotient of the probability of  $s$   $p_s$  and  $(1-p_s)$ .

This means that

- if a situation  $s$  is as likely as a situation  $t$ , the odds for  $s$  are  $0.5/(1-0.5)=1$ ;
- if a situation  $s$  is twice as likely as a situation  $t$ , the odds for  $s$  are  $0.667/(1-0.667)=2$ ;
- if a situation  $s$  is half as likely as a situation  $t$ , the odds for  $s$  are  $0.333/(1-0.333)=0.5$ .

In the present case, a value larger than 1 indicates that  $d$  becomes more likely, whereas a value smaller than 1 indicates that  $d$  becomes less likely. Thus, with regard to TRANSITIVITY: when the verb is transitive, the odds for the presence of  $d$  are significantly reduced (since  $or=0.203$  and, thus, smaller than 1). With regard to SENSE, the effects are described by seeing how the odds for the presence of  $d$  change when the verb sense is not 'spatial-reversive'. When the verb sense is 'customary', then the odds for the presence of  $d$  are reduced, but not significantly so. When the verb sense is 'reversive' or 'refractive', then the odds for the presence of  $d$  are significantly decreased; Table 3 provides the relevant statistics.

Factor	$\chi^2$	$df$	$p$	odds ratio $or$	lower $CI_{or}$	upper $CI_{or}$
TRANSITIVITY	16.56	1	<0.001	0.203	0.093	0.436
SENSE	28.24	3	<0.001			
SENSE: customary			0.44	0.39	0.048	8.48
SENSE: reversive			0.018	0.284	0.101	0.83
SENSE: refractive			<0.001	0.132	0.061	0.274

Table 3: Overview of the results of the binary logistic regression

### 3.2 Fine-grained exploration: hierarchical configural frequency analysis

We have already seen that both SENSE and TRANSITIVITY have significant effects on the absence or presence of  $d$ , and that the effect of SENSE is the stronger of the two. We now proceed with the HCFA to explore the effects with a higher degree of granularity. Consider Table 4 for an overview of all significant types and antitypes for the presence of  $d$  and Table 5 for an overview of all significant types and antitypes for the absence of  $d$ . The tables are interpreted by looking at the difference between the observed frequencies (cf. the column OBS) and the expected frequencies (cf. the column EXP); the directionality of the difference is indicated in the column RATIO column with ">" and "<" for types and antitypes respectively, and the column labeled  $Q$  provides the size of each effect, according to which the rows are sorted.

TRANSITIVITY	SENSE	OBS	EXP	RATIO	$p_{Holm}$	$Q$
intransitive	spatial-reversive	145	118.2	>	0.01	0.17
transitive		26	38.9	<	0.04	0.05

Table 4. Significant types and antitypes for the presence of  $d$

TRANSITIVITY	SENSE	OBS	EXP	RATIO	$p_{\text{Holm}}$	$Q$
intransitive	spatial-reversive	11	27.6	<	<0.001	0.07
	spatial-reversive	14	33.3	<	<0.001	0.08
intransitive		31	43.9	<	0.03	0.06
transitive	refactive	10	2.3	>	<0.001	0.03
transitive	reversive	8	0.9	>	<0.001	0.03
transitive		22	9.1	>	<0.001	0.05
	refactive	29	13.6	>	<0.001	0.06

Table 5. Significant types and antitypes for the absence of *d*

These results are interesting because they exemplify the merit of studying this phenomenon both multifactorially and on different levels of granularity. On the one hand, they are largely compatible with the results from the logistic regression: when the verb is transitive, the presence of *d* is dispreferred (row 2 of Table 4), the absence of *d* is preferred (row 6 of Table 5), and when the verb is intransitive, the absence of *d* is dispreferred (row 3 of Table 5).

On the other hand, the finer granularity of the HCFA indicates

- why the overall interaction SENSE  $\times$  TRANSITIVITY was only marginally significant;
- why SENSE came out as the variable more strongly correlated with MORPHEME;
- that there are a several configurations from this interaction that do reach significance.

The overall interaction is only marginally significant because, on the one hand, there are significant preferences such that intransitive spatial-reversives prefer the presence of *d* (and disprefer its absence), and both transitive refactives and transitive reversives prefer the absence of *d*. On the other hand, however, it also becomes clear that despite the significance of these

interactions, they do not contribute very much in terms of explanatory/predictive power. Consider, for example, the preference of intransitive  $\times$  spatial-reversives. This interaction has a small effect size ( $Q=0.07$ ) that is in fact slightly weaker than that of spatial-reversives alone ( $Q=0.08$ ) and slightly higher than that of intransitives alone ( $Q=0.06$ ). That is to say, in this case, adding to transitivity information about verb semantics increases predictive power (if only slightly) whereas adding to verb semantics information about transitivity adds only noise. Similarly, transitive  $\times$  refractive is a significant interaction type, but its strength ( $Q=0.03$ ) is again weaker than those of transitives or refractives in isolation ( $Q=0.05$  and  $Q=0.06$  respectively).

Before we proceed to discuss the implications of these quantitative results, let us point out one additional interesting aspect of the data. We have discussed several significant (anti)types from the interaction, but we also showed that these usually do not contribute much information beyond the significant main effects of SENSE and TRANSITIVITY. There is one noteworthy aspect of the overall interaction, though. Consider Figure 2, which represents the interaction in two Cohen-Friendly association plots. Light grey boxes indicate observed frequencies that are larger than expected, and dark grey boxes indicate observed frequencies that are less than expected; the sizes of the boxes reflect the difference between the observed and the expected frequencies (cf. Cohen 1980 or the documentation for *assocplot* in R for details).

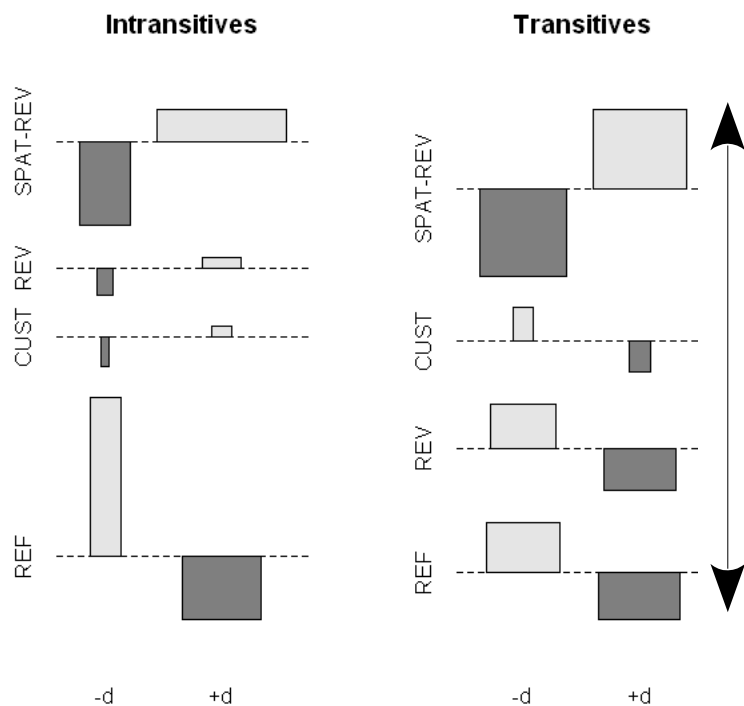


Figure 2: Association plots for *SENSE* × *MORPHEME*

The plot immediately reveals two effects that are not as easily identified from the statistics alone (and hardly at all by means of introspection). First, reflexives and spatial-reversives pattern identically in intransitives and transitives (dispreferring and preferring  $d$  respectively). However, customaries and reversives pattern slightly differently: in intransitives they pattern like spatial-reversives, but in transitives they pattern like reflexives. Even more interesting, however, is the second effect: in spite of the above difference, on the whole, the senses pattern such that they can be rank-ordered roughly in terms of their (dis)preference for  $d$  (hence the vertical arrow on the right): spatial reversives have the strongest preference for  $d$ , reflexives have the strongest dispreference for  $d$ , and customaries and reversives occupy the middle ground.

In the following section, we interpret these results with reference to what they tell us

about the distribution of *d* in Dena'ina iteratives as well as how that distribution compares to the behavior of Athabaskan middle constructions in general.

## **4. Discussion and conclusion**

### *4.1 Towards an explanation*

Our corpus-based quantitative study of Dena'ina iterative verbs tackles a question that has received contrasting explanations in the past. Some scholars adopted a predominantly syntactic perspective and explained middle-voice marking on the basis of intransitivity; others adopted a more semantic perspective. Our quantitative approach finds its origin in a previously offered motivation – the low differentiation of arguments and events – and we show empirically that (i) while there is some truth to both perspectives, (ii) neither alone tells the whole story. As to (i), it is true that intransitivity has an effect, but so does semantics: in fact the correlation of SENSE with *d* is significantly stronger than that of TRANSITIVITY. As to (ii), there is an interaction which, while only marginally significant in the coarse-grained analysis, in the fine-grained analysis gives rise to several significant types and antitypes. Thus, the heterogeneity in the literature is understandable: the picture is more complicated than a monofactorial approach can account for.

Let us now explore the findings from a more general perspective with a particular emphasis on first, the nature of the semantic cline illustrated in Figure 2 and second, what they tell us about Dena'ina iteratives in particular and middle constructions in general. Consider Figure 3, which organizes the four senses on different continua.

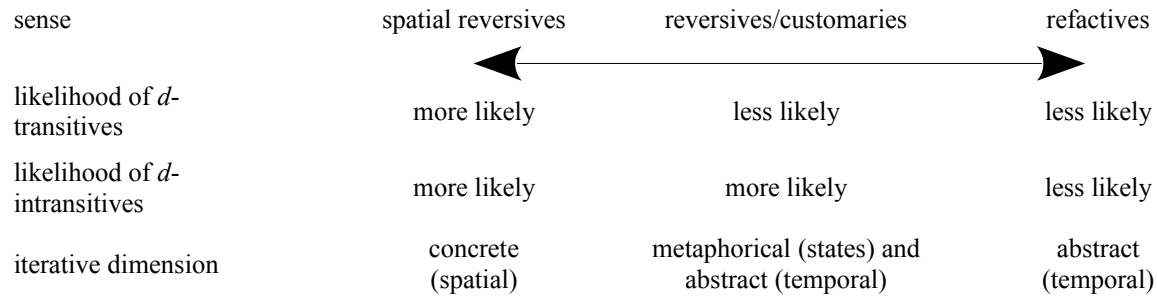


Figure 3: Continuum of senses of iterativity

Senses to the left are more likely to be middles, while senses to the right are less likely. While it is not surprising that the uses of the iterative morpheme fall into the semantic classes of reversives, spatial reversives, refactives and customaries – this is after all just how we coded the database – two interesting generalizations follow from the semantically-motivated patterning, specifically

- *d* is more preferred with the concrete, spatial senses (e.g. *nuidyu* 's/he is coming back' and *nughudnex* 's/he paddled back') and less preferred with the more abstract temporal senses (e.g., *nughilghatl* 'it got dark again' and *nuqeyltih* 'they used to wash him/her').
- while both intransitives and transitives exhibit nearly the same cline and ordering, they have a different cut-off point as to when *d* becomes preferred (cf. also again Figure 2).

There is a correlation between the concreteness/abstractness of senses and the presence/absence of *d*: the senses to the left are conceptually more concrete in their iterative dimension – that is, as to which part of the event is being repeated – and those to the right are conceptually more abstract. At the left, in spatial reversives, the iterative morpheme refers to

literal motion from a concrete source to a concrete goal (and back) – the repetition takes place in the spatial dimension. In our data, spatial reversives are by far most likely of all the senses investigated to be marked as middles. By returning to Rice's (2000 *a*) account of the unifying semantic condition of the suppression of differentiation of arguments and events, we can fairly easily reconcile these forms as middles. It is the nature of spatial reversives that the difference between the spatial starting and ending points of an action, i.e., the locational source and goal, is suppressed, much in the same way that the difference between initiator and recipient is suppressed in reflexive constructions.

To illustrate, consider the reflexive form *hudnil'an* 'he is looking at himself' and the spatial iterative form *nu'idyu* 'he came back'. In the former, the arguments are of low differentiation – indeed they have the same referent. But if we also conceive of them as the locations in space occupied by the referent, we can easily analogize to the spatial iterative. The source and goal occupy the same location, and the suppression of a difference between the two triggers middle marking. It is therefore consistent with previous accounts by Rice and others that, in Dena'ina iterative verbs, the spatial reversives are more likely than all other iteratives to be marked as middles with the *d*-classifier.

The intermediate position of the reversives in Figure 3 can be explained by considering that while they do not involve literal physical motion, they can be understood as metaphorically closely related to physical motion in terms of conceptual metaphor theory (cf. Lakoff and Johnson 1980; Lakoff and Turner 1989). Via the metaphor STATES ARE LOCATIONS (cf. Lakoff and Turner 1989: Ch. 2), changes from one state into another are conceptualized as movement from one place to another, and the return to a previous state is conceptualized as the return to a previous location. Thus, since the reversives are conceptually similar to spatial-

reversives, it is not surprising that their *d*-patterning differs only minimally from spatial-reversives.

Because of the small number of tokens of customaries, our explanation for their intermediate position is as yet admittedly more speculative. While customaries, like refactives (cf. below), involve temporal repetition, they differ from refactives by referring to habitual repetitions, assumedly more than one. We assume that repeated/habitual events are more likely to be lexicalized than just second occurrences of other individual events (it is simply more useful to speakers to have a separate lexical item for repeated/habitual occurrences of an event than for any event that happens to be repeated just once) but also more likely to be lexicalized as nouns. For instance, it has been noted in Cognitive Grammar (cf., e.g., Langacker 1991:34f.), that nominalization (among other things) involves conceptual reification or the conceptualization of the distinct stages of a process/activity as a single separate episode.

Once a habitual action becomes nominalized this way, it is categorized as a member of the notionally-defined class of THINGS and as such becomes construed as more similar to the more typical members of the class of THINGS, namely concrete objects (cf. Langacker 1987: Ch. 5), and the intermediate position of customaries thus conceptualized between concrete locations and the completely abstract refactives would be expected.<sup>9</sup>

Finally, at the right of the abstract/concrete continuum are the refactives, which are least likely of all iterative senses to be marked as middles. These refer to a second/subsequent occurrence of an action – in these, the repetition takes place in the abstract temporal dimension.

In summary, concrete entities – the locations of the source and goal of an action – qualify most for suppression: regardless of transitivity, the difference *between two locations* can be suppressed to such a degree that middle marking is triggered, but the temporal difference

*between actions repeated in time* resists suppression. However, intransitivity is still important: in the case of intermediately concrete iterativity, there is the slight interaction with TRANSITIVITY such that the degrees to which reversives and customaries prefer *d* is dependent on that variable.

We can now extend our account of the semantic motivation for Dena'ina middles to include the suppression of the differentiation of arguments, events, *and locations*. In spatial reversives – as in reflexives, reciprocals, self-benefactives, and other middle constructions – two entities merge to form a single referent (cf. Rice 2008). We now have an explanation for the presence of *d* that unifies iteratives with other middle constructions, and we can update our Venn diagram to Figure 4.

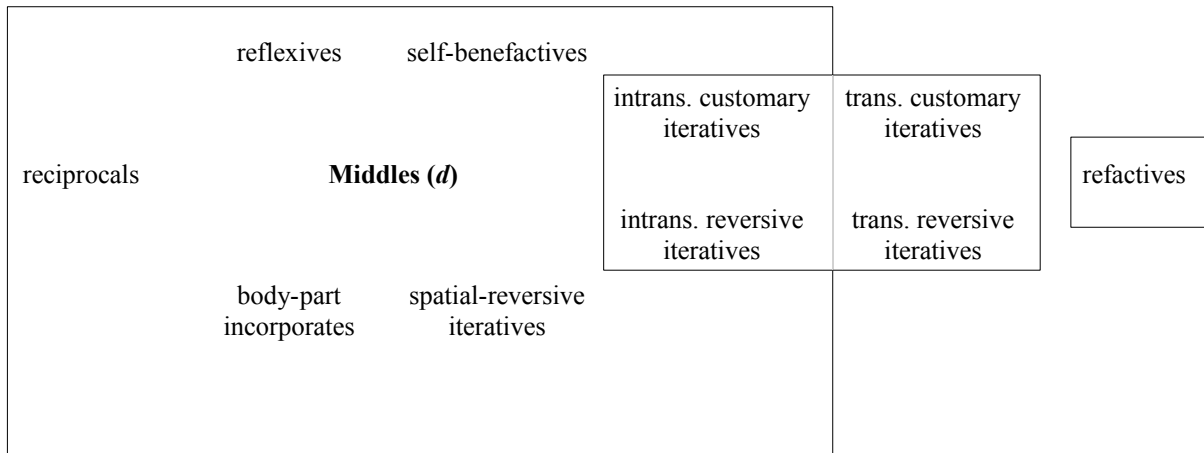


Figure 4: Inclusion of various constructions in a formal class of middles (updated)

The analysis presented here allows Dena'ina iteratives to be unified with other types of middles. Rice (2000 *a*) provides a summary of fourteen types of Athabaskan middle constructions, eight of which can be accounted for by the suppression of the differentiation of arguments and events. By building on and refining her account to include the suppression of the differentiation of

locations, we can now add a ninth type, for Dena'ina at least, to the list of middle constructions with unified semantic motivation.

#### 4.2 *Concluding methodological remarks*

On a more methodological note, we would like to emphasize the role the quantitative and corpus-based methodology played in this paper. Such methods are currently and rapidly becoming increasingly popular both in linguistics in general and linguistic typology and language description in particular (cf. Bickel et. al. 2006 and Cysouw, to appear, for just two other successful corpus-based approaches to lesser-studied languages). Scholars are increasingly realizing that such methods have much to offer in terms of interesting generalizations that would not be otherwise apparent, and that they need not only be applied to the large frequencies possible for languages for which multi-million word corpora are available.

Even a corpus as small as the one studied here yielded findings that both conform to previous work well enough to show that the overall approach is meaningful and also allow us to arrive at distributional patterns, generalizations, and conclusions that would be difficult to obtain otherwise. It is hard to imagine how one could detect the interactions present in the Dena'ina iteratives data in any other way. It is our hope, therefore, that, apart from its theoretical points, this paper will encourage further use of quantitative methods in the study of similarly underrepresented languages. An object of study as complex as human language exhibits patterns that are not always visible to the naked eye, and at such a level of complexity, the kind of approach exemplified here can help us reach further with our explanations.

## References

- Arce-Aranales, Manuel; Melissa Axelrod; and Barbara A. Fox. 1994. Active voice and middle diathesis: A cross-linguistic perspective. *Voice: Form and function*, ed. Barbara A. Fox and Paul J. Hopper, pp. 1-21. Amsterdam, Philadelphia: John Benjamins.
- Bickel, Balthasar; Dirk Janssen; and Fernando Zúñiga. 2006. Randomization tests in language typology. *Linguistic Typology* 10: 19-40.
- Cohen, A. 1980. On the graphical display of the significant components in a two-way contingency table. *Communications in Statistics – Theory and Methods*, A9, 1025-41.
- Collins, James. 1979. Athapaskan classifiers, person and deference. *Berkeley Linguistic Society* 5: 49-62.
- Cook, Eung-Do. 1984. *A Sarcee grammar*. Vancouver: University of British Columbia Press.
- Cusic, David. 1981. *Verbal plurality and aspect*. Ph.D. dissertation, Stanford University.
- Cysouw, Michael. To appear. Quantitative explorations of the world-wide distribution of rare characteristics, or: The exceptionality of north-western European languages. *Exception in language*, ed. Horst Simon and Heike Wiese. Berlin, New York: Mouton de Gruyter.
- von Eye, Alexander. 1990. *Introduction to configural frequency analysis: the search for types and antitypes in cross-classifications*. Cambridge: Cambridge University Press.
- Gries, Stefan Th. 2004. *HCFA 3.2: a program for R for Windows*.
- Goddard, Pliny Earle. 1905. *The morphology of the Hupa language*. University of California Publications, vol. 3. Berkeley: University of California Press.
- Hoijer, Harry. 1946. The Apachean verb, part III: The classifiers. *IJAL* 12: 51-59
- Holton, Gary. 2000. *The phonology and morphology of the Tanacross Athabaskan language*. Ph.D. dissertation, University of California, Santa Barbara.

- , 2005. Agent affectedness and the Athabaskan middle voice. Paper presented at Eastern Michigan University Linguistics Colloquium, Ypsilanti.
- Hopper, Paul J. and Sandra A. Thompson. 1980. Transitivity in discourse and grammar. *Language* 56: 251-99.
- Howren, Robert. 1971. A formalization of the Athapaskan d-effect. *IJAL* 37: 96-113.
- Jetté, Jules and Eliza Jones. 2000. *Koyukon Athabascan dictionary*. Fairbanks: Alaska Native Language Center.
- Kari, James. 1979. Athabaskan verb theme categories: Ahtna. Alaska Native Language Center Research Papers, no. 2. Fairbanks: Alaska Native Language Center.
- , 1990. *Ahtna Athabascan dictionary*. Fairbanks: Alaska Native Language Center.
- , n.d. *Dena'ina stem dictionary (draft)*. Lexware file.
- Kemmer, Suzanne. 1993. *The middle voice*. Amsterdam, Philadelphia: John Benjamins.
- Kibrik, Andrej A. 1993. Transitivity increase in Athabaskan languages. *Causatives and transitivity*, ed. Bernard Comrie and Maria Polinsky, pp. 47-67. Amsterdam: John Benjamins.
- , 1996. Transitivity decrease in Navajo and Athabaskan: Actor affecting propositional derivations. *Athabaskan language studies: Essays in honor of Robert W. Young*, ed. Eloise Jelinek; Sally Midgette; Keren Rice; and Leslie Saxon, pp. 259-303. Albuquerque: University of New Mexico Press.
- , 2002. Reflexivization and detransitivization processes in Athabaskan languages. Ms., Institute of Linguistics, Russian Academy of Sciences, Moscow.
- Krauss, Michael. 1969. On the classification in the Athapaskan, Eyak and Tlingit verb. *Indiana University publications in anthropology and linguistics* 24. Baltimore: Waverly Press.

- Lakoff, George and Mark Johnson. 1980. *Metaphors we live by*. Chicago: University of Chicago Press.
- Lakoff, George and Mark Turner. 1989. *More than cool reason: a field guide to poetic metaphor*. Chicago: University of Chicago Press.
- Langacker, Ronald. 1987. *Foundations of cognitive grammar, volume I*. Stanford: Stanford University Press.
- , 1991. *Foundations of cognitive grammar, volume II*. Stanford: Stanford University Press.
- Leer, Jeff. 1979. Proto-Athabaskan verb stem variation. *Alaska Native Language Center Research Papers*, no. 1. Fairbanks: Alaska Native Language Center.
- McDonough, Joyce. 1989. Argument structure and the Athapaskan "classifier" prefix. *Proceedings of the West Coast Conference on Formal Linguistics* 8: 220-45.
- R Development Core Team. 2008. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <<http://www.R-project.org>>.
- Rice, Keren. 1989. *A grammar of Slave*. Berlin: Mouton de Gruyter.
- , 2000a. Voice and valency in the Athapaskan family. *Changing valency: Case studies in transitivity*, ed. R.M.W. Dixon and Alexandra Y. Aikhenvald, pp. 172-235. Cambridge: Cambridge University Press.
- , 2000b. Morpheme order and semantic scope: Word formation in the Athapaskan verb. Cambridge: Cambridge University Press.
- , 2008. What's in a middle? Paper presented at the 11<sup>th</sup> Annual Workshop on American Indigenous Languages, Santa Barbara.
- Shluinsky, Andrey. 2007. Universal gram-types in the domain of pluractionality. Paper presented

- at the Association for Linguistic Typology 7<sup>th</sup> biennial meeting, Paris.
- Tenenbaum, Joan. 1976a. Dena'ina Sukdu'a I/Tanaina stories I. Fairbanks: Alaska Native Language Center.
- 1976b. Dena'ina Sukdu'a II: Chulyin sukdu'a/Tanaina stories II: Raven stories. Fairbanks: Alaska Native Language Center.
- 1976c. Dena'ina Sukdu'a III: Dghiliq' sukdu'a/Tanaina stories III: Mountain stories. Fairbanks: Alaska Native Language Center.
- 1976d. Dena'ina Sukdu'a IV: Nanutset k'ughun nił t'qul'an/Tanaina stories IV: Stories of the wars they had before our times. Fairbanks: Alaska Native Language Center.
- 1978. Morphology and semantics of the Tanaina verb. Ph.D. dissertation, Columbia University.
- Thompson, Chad. 1989. Voice and obviation in Athabaskan and other languages. Ph.D. dissertation, University of Oregon.
- 1996. The Na-Dene middle voice: An impersonal source of the D- element. *IJAL* 62: 351-78.
- Tuttle, Siri G. 2008. Phonetics and word definition in Ahtna Athabaskan. *Linguistics* 42: 439-70.
- Wood, Esther. 2007. The semantic typology of pluractionality. Ph.D. dissertation, University of California, Berkeley.
- Xrakovskij, Viktor S. (ed.). 1997. Typology of iterative constructions. München, Newcastle: LINCOM EUROPA.

1 We wish to thank Anthony Aristar, Gary Holton, Ljiljana Progovac, Martha Ratliff, Siri Tuttle  
and two anonymous IJAL reviewers for their comments on earlier versions of this paper. We  
especially thank Jim Kari, Andrej Kibrik and Keren Rice for their extensive advice and detailed  
suggestions; the authors are fully responsible for any errors. Gloss abbreviations: ASP = aspect;  
CLF = classifier; CONJ = conjugation; GEND = gender; INCEP = inceptive; INCORP =  
incorporated noun; ITER = iterative; PERF = perfective; SUB = subject; TERM = terminal;  
OBJ = object; 1SG:D = portmanteau of 1s. subject and *d*-classifier. Source abbreviations: TDS  
*a-d* = Tenenbaum (1976 *a-d*).

2 When *nu* follows a syllable with /a/ in the nucleus, the vowel of the iterative harmonizes to  
form /na/; cf. *qana'ilghel* 'he lay back down'.

3 Other labels have been suggested for the unfortunately-named *classifier* morpheme, including  
*transitivity indicator* (Kibrik 1993; 1996; 2000) and *voice/valence* (Rice 2000 *a*; 2000 *b*). We  
continue to use the traditional term here. See Kibrik (1996) for the history of the misnomer.

4 McDonough (1989) argues that *d* does in fact eliminate an argument, causing a change to  
argument structure.

5 We are grateful to Andrej Kibrik for these suggestions for the categories of SENSE.

6 Tuttle (2008) describes the assumptions present in verb theme entries in Athabaskan  
dictionaries like Kari (n.d.). She writes “[v]erb theme entries are a formal device for showing  
what elements have to be listed in a dictionary for a person who knows the language to be able  
to reconstitute the verb [...] Verb themes state the lexical requirements and the argument  
structure of Athabaskan verbs” (441f.). It is by this logic – the same logic that was applied  
during the creation of the dictionary – that we can determine what constitutes underlying  
material and what constitutes derivational material.

7 The accuracy of the database was independently verified by James Kari, to whom we are  
grateful.

- 8 All statistical computations and plots were done with the software environment for statistical computing and graphics R (for Windows) version 2.6.2 (cf. R Development Core Team 2008). For the HCFA, we used the program HCFA 3.2 (Gries 2004), which uses Holm's correction for multiple testing for individual tests to avoid inflating the probability of type-I errors.
- 9 While we think this is a possible approach, the positioning of customaries on this cline does not pose a problem to the overall generalization even if our speculation is not borne out. Even without the assumption of repetition-supports-reification, the position of customaries fits perfectly in the case of intransitives – customaries refer to temporal repetition much like reflexives and in intransitives these two are grouped together – and rather well in the case of transitives. We consider the position of customaries in transitives to fit "rather well" because (i) it is not optimal since customaries are not grouped together with the reflexives like a perfect semantically-based patterning as with the intransitives would lead us to expect, but (ii) the customaries neither are in the worst position they could theoretically occupy on this cline (i.e., above spatial reversives) nor is their statistical patterning anywhere close to that of spatial reversives: contrary to those, customaries still disprefer *d*.