

On the distinction between regular and irregular inflectional morphology:  
evidence from Dinka

D. Robert Ladd, Bert Remijsen, Caguor Adong Manyang\*  
University of Edinburgh and \*University of Bahr el-Ghazal

To appear as a Short Research Report in *Language*

***Abstract***

Discussions of the psycholinguistic significance of regularity in inflectional morphology generally deal with languages in which regular forms can be clearly identified and revolve around whether there are distinct processing mechanisms for regular and irregular forms. We present a detailed description of Dinka's notoriously irregular noun number inflection and suggest that no pattern can usefully be designated as regular. Psycholinguistic studies of Dinka would make a valuable contribution to our understanding of inflectional morphology cross-linguistically.\*

\* This research was sponsored primarily by the U.K. Arts and Humanities Research Council (Research Grant no. MRG-AN11781/APN19394 to Edinburgh University). The involvement of Caguor Adong Manyang was made possible by a Small Research Grant (no. 39265) from the British Academy. The research also forms part of Ladd's project 'Simultaneous and sequential structure in language', supported by a Leverhulme Individual Research Fellowship. We gratefully acknowledge this support. An earlier version of this paper was presented at the 15th Manchester Phonology Meeting in 2007.

## ***1. Introduction.***

Discussion of the acquisition and processing of inflectional morphology has for some time been dominated by a debate over the relation between regular and irregular forms. According to the ‘dual-route’ hypothesis (e.g. Pinker 1999, Pinker and Ullman 2002), there are two separate processing mechanisms, one based on the productive application of rules and the other based on lookup of stored forms in the mental lexicon. There are stronger and weaker versions of this view, but all presuppose that regular forms are normally parsed into, and assembled from, their component parts, while irregular forms are normally stored whole in the lexicon. The opposing view, based originally on connectionist modeling (beginning with Rumelhart and McClelland 1986; for more current reviews see McClelland and Patterson 2002 and Baayen 2007), holds that all morphological acquisition and processing is based on the development of associations between morphologically related forms, and that the strength of the associations depends on a large number of factors, including the frequency of the forms and the frequency of the patterns involved. In between these two extremes lie a range of empirical studies of morphological acquisition and processing, which argue in various ways that regularity is a matter of degree. For example, the work of Ravid and her colleagues (e.g. Laaha et al. 2006 on noun plurals in German, Ravid and Schiff 2009 on noun plurals in Hebrew) is based on a two-dimensional characterization of regularity, involving degrees of ‘suffix regularity’ and ‘stem change’, and shows that such a description is consistent with many findings about how morphology is acquired. Albright’s work (2002) on sub-regularities in Italian verb morphology suggests that there are ‘islands of reliability’ within the set of irregular forms in any given system, and that speakers are able to make productive use of these sub-regularities.

What all this work shares, however – in addition to being based largely on well-studied languages – is the assumption that it is possible to distinguish regular and irregular (or more regular and less regular) inflection. Even work that explicitly argues for a gradient conception of regularity (like that of Ravid and her colleagues), or connectionist work that seeks to explain regularity without any special mechanisms, takes the notion of regularity itself for granted. In this squib we present data from the noun number marking system of Dinka, which make it appear entirely possible for a rich inflectional system not to have any patterns that can be identified as regular at all. We briefly consider the potential implications of this for theories of morphological acquisition and processing.

### **1.1. Background information on Dinka.**

Dinka is a Western Nilotic language within the Nilo-Saharan family, spoken by more than two million speakers (Gordon 2005). The Dinka homeland is along the banks of the White Nile and its tributaries in Southern Sudan, but there are now large communities in Khartoum and in other countries as well. A description of the Dinka sound system (Agar dialect) can be found in Andersen (1987). Dinka is of particular phonological interest because of its rich suprasegmental system, which includes independent distinctions of tone, vowel length and voice quality. The dialect represented here (Luanyjang, part of the Rek dialect cluster) has four tonemes (Remijsen and Ladd 2008), three degrees of vowel length (Remijsen and Gilley 2008), and a two-way distinction of voice quality (breathy vs. modal or creaky) (Remijsen and Manyang 2009); the details of other dialects (many of which are essentially undocumented) differ in detail but are broadly similar.

Both the verb and the noun systems of Dinka involve rich inflectional paradigms (cf. Andersen 1993 and 2002, respectively). For example, Dinka transitive verbs appear in a range of inflected forms, marking tense, person, number, valence (active, passive, non-topical subject, intransitive), direction (centripetal, centrifugal), and modality (negation). Illustrations are given in Table 1. As can be seen from these illustrations, verb inflections are marked primarily by stem-internal changes, affecting length, tone, vowel height, and voice quality. In contrast to the situation familiar from English and many other languages (Bybee and Newman 1995), these stem-internal inflections are regular within classes. For example, 1<sup>st</sup> singular formation invariably involves vowel lowering in transitive verb stems, unless the stem vowel is /a/, which cannot lower further.

Finite (topical subject)	-l̥eɛl	Finite (topical subj., intrans.)	-l̥ɛɛl
Finite (non-topical subj.)	-léeeɛl	Finite passive	-l̥eɛl
Finite 1 <sup>st</sup> singular	-l̥eɛɛl	Inf. past (topical subj.)	léeeɛl
Finite 2 <sup>nd</sup> singular	-léel	Inf. past (non-topical subj.)	l̥eɛl
Finite 3 <sup>rd</sup> singular	-l̥eɛl	Inf. passive past	léeeɛl
Finite 1 <sup>st</sup> plural	-l̥eɛl-k̥ú	Inf. negation	l̥eɛl
Finite 2 <sup>nd</sup> plural	-l̥eɛl-k̥í	Inf. passive past	léeeɛl
Finite 3 <sup>rd</sup> plural	-l̥eɛl-k̥í	Deverbal noun	l̥ɛɛl

Table 1. *Inflections of a transitive Dinka verb, illustrated by /leel/ ‘to challenge’.*

Dinka’s stem-internal inflectional system has developed out of a suffix-based system, which can still be witnessed in more conservative sister languages such as Pári (Andersen 1990). As a result of processes such as compensatory lengthening and the reassociation of tones, the marking of inflectional distinctions in Dinka has been transferred from suffixes to the stem. The relation between the original suffixes and stem-internal marking is illustrated in 1, from Andersen (1990:16-17).<sup>1</sup>

(1)	Schematic representations of quantity		Examples		Gloss
	Pári	Dinka	Pári	Dinka	
	CVC	CVC	p̥ɪ̃p̥	p̥ɪ̃p̥	‘land’
	CVC-V	CVVC	c̥ɪ̃ŋ-ò	c̥ĩin	‘hand’
	CVVC	CVVC	ɲàaŋ	ɲáaŋ	‘crocodile’
	CVVC-V	CVVVC	c̥ĩin-ò	c̥ĩiin	‘intestine’

The loss of suffixes seems to have driven the development of Dinka’s three-way distinction of vowel length (Remijsen and Gilley 2008). Also as a result of the loss of suffixes, Dinka words are predominantly monosyllabic.

## 1.2. Number marking on Dinka nouns.

The Nilo-Saharan languages are well known for the complexity of their nominal morphology (Welmers 1973:23, Dimmendaal 2000). Among other things, number marking is semantically more complex than in many other languages, in that singular is not always unmarked. Nouns

<sup>1</sup> While the Pári examples come from Andersen (1990), the transcriptions for the Dinka forms come from our own data, representing the Luanyjang dialect on which our own study is based.

referring to things that are typically encountered in masses, sets or pairs (e.g. grass, ants, fingers, or eyes) are often unmarked in the plural, and in many Nilo-Saharan languages are marked by an affix in the singular. Nouns referring to things that are typically encountered as countable individuals (e.g. chief, river, cattle camp) are often unmarked in the singular, and appear with an affix in the plural. The semantic basis of the system is thus related to the distinction between mass and count nouns found in many other languages. However, as with noun class systems everywhere, it is not always easy to detect a semantic basis for the morphological treatment of any given noun; moreover, there is a third group of nouns that has affixes for number in both the singular and the plural. This ‘tripartite’ number marking system is found across the Nilo-Saharan language family (Dimmendaal 2000); this is illustrated in 2 with examples from Anywa taken from Reh (1999), cited in Storch (2005:226ff):

(2) Tripartite number marking in Anywa

Singular	Plural	Gloss	
cèn-ò	cɛŋ	‘hand’	(marked singular)
èc	íqí	‘belly’	(marked plural)
túú-ó	tùù-è	‘palm tree’	(both singular and plural marked)

In Dinka, number inflection involves the tripartite system just outlined to at least some extent; in addition, nouns are inflected for locative case, and to signal the presence of a modifier (Andersen 2002). As explained in the previous section, however, Dinka has lost most of its suffixes, and number marking that is signaled by an affix in a typical Nilo-Saharan language often involves a longer vowel in Dinka. The corresponding words in Dinka to the Anywa examples just given are shown in 3:

(3) Tripartite number marking in Dinka

Singular	Plural	Gloss	
cíin	cìn	‘hand’	(marked singular)
jééc	jèeec	‘belly’	(marked plural)
tùuk	túuk	‘fruit of palm’	(both singular and plural marked)

More generally, number marking in Dinka involves the same stem-internal processes found in the verb morphology, namely changes in tone, vowel quality, voice quality and the nature of the coda consonant. Unlike the verb morphology, though, where the patterns of modification are quite fixed (as in the example of 1<sup>st</sup> singular formation above), no obvious generalizations hold for number marking on nouns. The only generalization with any degree of validity is the one involving vowel length just described, which seems to hold across a substantial proportion of nouns.<sup>2</sup> However, none of the other stem-internal phonological dimensions involved in marking number appear to show any such predictability.

The range of variation is illustrated in 4. Number can be marked by the change of a single phonological parameter, e.g. tone in 4a or vowel length in 4b. In 4c,d, the difference between singular and plural is marked by changes in both vowel length and tone – but different

<sup>2</sup> It is important, though, not to exaggerate the extent of the semantic regularity. Among animal names, for example, there are cases in which names of animals typically encountered in herds or groups show a longer vowel in the plural (e.g. /kèew / kèeet/ ‘[kind of] gazelle SG/PL’) and names of typically solitary animals show a longer vowel in the singular (e.g. /kjéet / kí/ ‘scorpion SG/PL’).

changes in the two cases. In 4e,f, three parameters are involved – tone, vowel height, and breaking or diphthongization in 4e, and tone, vowel length and voice quality in 4f. Finally, the examples in 4g,h illustrate cases involving four changes: vowel length, tone, vowel height, and voice quality in 4g, and tone, vowel height, voice quality and the coda consonant in 4h.

(4)	Singular	Plural	Gloss
a.	p̣iic	p̣iic	‘stirring stick’
b.	ḅiɲ	ḅiɲɲ	‘cup’
c.	ḅaɲ	ḅaɲɲ	‘chief’
d.	nòoon	nóon	‘grass’
e.	ṭiil	ṭj̣èɛl	‘thistle’
f.	wǎal	wàl	‘plant’
g.	lwák	lwěek	‘cattle byre’
h.	róow	ròoṭ	‘hippo’

The examples in 4 are not isolated exceptions, but are typical of Dinka number inflection, suggesting that there may be no regular way to mark number in Dinka at all – in other words, no group or class of nouns that can be treated as ‘regular’. This squib reports a detailed investigation of this possibility. We aimed to answer two specific questions: First, what combinations of phonological parameters are involved in the marking of number in Dinka nouns, and how many such combinations are there? Second, are there any obvious sub-regularities or conditioning factors determining which combination is used in any given case, and is there any evidence for a default choice?

## 2. *The study.*

### 2.1. Dataset.

For 400 Dinka nouns, we collected both the singular and the plural forms. This is not an entirely trivial task, in part because of the phonetic nature of the distinctions involved. The three-way length distinction can be difficult to hear in isolation; one of the tonal distinctions is neutralized or nearly neutralized in citation form (Remijsen and Ladd 2008); there are allophonic interactions of voice quality and vowel quality (Remijsen and Manyang 2009); and in the still-developing orthography the suprasegmental distinctions are represented inadequately or not at all. Consequently we took great care to be sure that the forms we report are correct, putting words in different sentence contexts and carrying out instrumental acoustic measurements where necessary, and we are confident that the data can be considered reliable. As noted above, the dialect under study is Luanyjang Dinka, the sound system of which is familiar to us from several studies carried out in conjunction with the work reported here (Remijsen and Gilley 2008; Remijsen and Ladd 2008; Remijsen and Manyang 2009). The data are based on the speech of the third author, but have been checked for consistency with other speakers in Khartoum; for only 10 nouns out of the 400 was there any variability.

The set of nouns investigated includes (a) 312 native monosyllabic words<sup>3</sup>, which are characteristic of the great majority of the Dinka lexicon; (b) 61 native polysyllabic but

<sup>3</sup> More accurately, the native monosyllables under investigation include 312 singular-plural pairs. The sample contains eight singulars (words meaning ‘child’, ‘boy’, ‘girl’, ‘woman’, ‘man’, ‘cow’, ‘bull’, and ‘goat’) that have two semantically distinct plurals, one referring to countable individuals (here abbreviated PL) and one referring to a collective (here abbreviated UNC); these were counted as 16 distinct singular-plural pairs rather than eight lexical items.

monomorphemic words, which invariably begin with /a-/<sup>4</sup>; (c) 17 examples of nominalizations and other morphologically complex nouns; and (d) 10 fairly recent loan words. For each of these nouns we coded the phonological parameters that distinguish singular from plural. For example, for /lwák/ vs. /lwëék/ ‘cattle byre:SG/PL’, we report differences in vowel height, vowel length, voice quality, and tone. Altogether we recorded differences in seven phonological parameters. For four of these parameters (vowel height, vowel backness, voice quality, and vowel breaking<sup>5</sup>), we recorded only whether the parameter differs or not between singular and plural.<sup>6</sup> For the other three parameters (vowel length, tone, and coda consonant), we recorded specific types of difference. For vowel length, we recorded whether there was a change of one degree (between short /V/ and medium /VV/ or between medium /VV/ and long /VVV/) or two (between short /V/ and long /VVV/); for tone we recorded the identity of the two tonemes involved, since all six possible pairs of tonemes are represented in the range of tone changes we recorded; for coda consonant we distinguish between alternations involving /t̩/ and all other alternations. The complete coding scheme is as follows:

(5)	<i>Parameter</i>	<i>Degrees or types</i>
	Vowel length	3 (same, different by 1 degree, different by 2 degrees) <sup>7</sup>
	Tone	7 (same, L~H, L~HL, L~LH, H~HL, H~LH, HL~LH)
	Coda	3 (same, C~ t̩, other difference)
	Vowel height	2 (same, different)
	Vowel backness	2 (same, different)
	Breaking	2 (same, different)
	Voice quality	2 (same, different)

Note that in the first instance we are concerned only with differences between singular and plural; we make no assumption about which form is unmarked and which is marked, because of the general nature of Nilo-Saharan number morphology noted earlier. That is, our coding is neutral with regard to the direction of the difference, and e.g. a difference in vowel length between short /V/ and medium /VV/ is coded as a difference of one degree of vowel length regardless of which vowel length occurs in the singular form and which in the plural. However, we return to this point in section 2.4 below.

<sup>4</sup> Counted as monosyllabic rather than polysyllabic are two words involving reduplication, one in the plural only (/tj̩ëék/ vs. /tj̩ëektj̩ëék/ ‘marriage:SG/PL’) and one in both singular and plural (/jakj̩ak/ vs. /j̩akj̩ak/ ‘lung:SG/PL’).

<sup>5</sup> We use the term ‘breaking’ to refer to an alternation in which one of the forms has a semivowel /j/ or /w/ between the onset consonant and the vowel, as in 4e above.

<sup>6</sup> In treating vowel height in this way, we ignore the extent of the difference in height, e.g. one degree, as in /i/ to /e/, or two degrees, as in /i/ to /ɛ/. In this we follow Andersen (1993, 2002), who suggests that the number of degrees of vowel height involved in the change is predictable from the absence (one degree) or presence (two degrees) of vowel breaking. Our dataset actually shows a few exceptions to this generalization, but they are so few that they did not seem to warrant complicating our data coding.

<sup>7</sup> In one segmental context, namely when the coda is /r/, medium (/VV/) vowels do not occur. Alternations between short (/V/) and long (/VVV/) in this context were coded as involving a single degree of vowel length. For further detail on vowel length before coda /r/ see Remijsen and Gilley 2008, Remijsen and Manyang 2009.

## 2.2. Native monosyllabic nouns.

Here we consider 307 native monosyllabic nouns.<sup>8</sup> Within this set, we found 81 different combinations of the above-mentioned parameters. There is great variation in the extent to which the various parameters are involved: a difference in Tone is involved in the marking of number for 85% of these nouns, whereas vowel Backness differs only for 1%. The values for the other parameters lie between these extremes: 71% for vowel Length; 49% for vowel Height, 21% for vowel Breaking, and 11% for both Voice quality and Coda consonant. The ten most frequent combinations of parameter differences are shown in Table 2.

1. Length (1), Tone (L~H)	12.0% (37)
2. Tone (L~H)	6.8% (21)
3. Length (1), Tone (L~LH)	6.5% (20)
4. Height, Length (1), Tone (L~H), Breaking	6.2% (19)
5. Length (1), Tone (L~HL)	5.2% (16)
6. Height, Length (1), Tone (L~H)	4.2% (13)
7. Height	3.6% (11)
8. Height, Tone (L~H), Breaking	3.2% (10)
9. Height, Length (1), Tone (L~LH)	3.2% (10)
10. Length (1), Tone (H~LH)	2.9% (9)

Table 2: The most common patterns of number marking for native monosyllabic nouns in Luanyjang Dinka, with percentage (%) and number of cases (in brackets).

As seen from Table 2, combinations of Length, Tone and Height differences are particularly common. These ten most frequent combinations account for 57% of the set of native monosyllabic nouns. The patterns involved in the marking of number in the remainder of these nouns are represented by between 1 and 8 cases.

## 2.3. Native polysyllabic nouns.

Among native stems, the pattern /a(CV)CV(V)(V)C/ is the only polysyllabic noun template (cf. Andersen 1987, Storch 2005:168). The initial vowel may have a derivational origin, but synchronically it is part of the stem, and the words can be considered monomorphemic. Tone is specified only on the final syllable (Remijsen and Ladd 2008: 195-196) and only the final syllable is involved in stem-internal inflectional changes. Three examples are presented in 6.

(6)	Singular	Plural	
a.	apwóol	apól	‘maize’
b.	ap̣iḳòol	ap̣iḳòool	‘story’
c.	atjêep	aṭiip	‘shelter’

Our dataset includes 61 such nouns. Within this set, there are 33 combinations of phonological parameter differences. All but one of these 33 combinations also occurred in the

<sup>8</sup> Of the original set of 312, we exclude from consideration five that may be said to involve suppletion, in which singular and plural differ in onset consonant as well as in other ways, namely /tiik/ vs. /djàaar/ ‘woman, wife:SG/PL’, /tiik/ vs. /djòoor/ ‘woman, wife:SG/UNC’, /wéenj/ vs. /tuq̣òok/ ‘cow:SG/PL’, /ràaan/ vs. /kóoc/ ‘person:SG/PL’ and /mòoc/ vs. /ròoor/ ‘man: SG/PL’. For the two plural forms of ‘woman, wife’ see footnote 3.

set of native monosyllabic set of nouns. The ten most frequent combinations of parameters are listed in Table 3.

1. Length (1), Tone (L~H)	9.8%	(6)
2. Length (1), Tone (L~LH)	8.2%	(5)
3. Length (1), Tone (H~LH)	8.2%	(5)
4. Length (1), Tone (L~HL)	6.6%	(4)
5. Length (1), Tone (LH~HL)	6.6%	(4)
6. Length (2), Tone (L~H)	4.9%	(3)
7. Height, Length (1), Tone (L~H)	4.9%	(3)
8. Length (1)	3.3%	(2)
9. Length (1), Tone (H~HL)	3.3%	(2)
10. Length (2), Tone (L~HL)	3.3%	(2)

*Table 3: The most common patterns of number marking for native polysyllabic nouns in Luanyjang Dinka, with percentage (%) and number of cases (in brackets).*

These ten patterns cover 59% of the native polysyllabic nouns. Impressionistically comparing this top-ten with the one for monosyllabic nouns, we find that combinations of vowel Length and Tone are more prevalent among the set of polysyllabic nouns, suggesting that vowel Height plays less of a role here than in the monosyllabic set. Indeed, a count reveals that differences of vowel Height are involved in number marking in only 39% of the polysyllabic nouns, as compared to 49% for the monosyllables. In other respects the monosyllabic and polysyllabic nouns seem to behave similarly.

#### **2.4. Sub-regularities in native monomorphemic noun number marking.**

The fact that vowel Height differences are used in number marking to a greater extent in monosyllables than in polysyllables could be seen as a sub-regularity in the number marking system, of the sort investigated for Italian verb conjugations by Albright (2002). That is, knowing that vowel Height differences are less common in native polysyllabic nouns could provide at least probabilistic support for the learner of Dinka in reliably storing singular-plural pairs. In this section we summarise our observations on such sub-regularities and ‘islands of reliability’ seen in the parts of our sample discussed so far. There clearly are some such generalizations, especially with regard to Tone, though none of them go beyond probabilistic statements, and many are restricted to very specific subsets of the corpus. Except where explicitly noted, however, all these generalizations appear to apply similarly in both monosyllables and polysyllables.

- i. With regard to Length, roughly 60% of nouns differing by one degree of Length (249 cases) involve a medium (/VV/) and a long (/VVV/) vowel, while only 40% involve a short (/V/) and a medium (/VV/) vowel. The overwhelming majority (>80%) of nouns that do *not* differ in Length (99 cases) are medium (/VV/) in both singular and plural, and are almost never short (/V/) (only 3 cases). Where differences of Length are involved in number marking, they are about equally likely to involve lengthening from singular to plural as from plural to singular (cf. the comments on number marking in section 1.2).
- ii. With regard to Height, nouns are less likely to have number marked by a difference of Height if it is also marked by a difference of Length. Some 65% of singular-plural pairs that do not differ in Length differ in Height, whereas only 40% of pairs that differ in Length also differ in Height. Where there are differences of Height, they are



about equally likely to involve lowering from singular to plural as from plural to singular (again, cf. the comments on number marking in section 1.2).

- iii. In nouns where number is marked by differences of both Length and Height (106 cases), the lower Height tends to co-occur with the greater degree of Length, regardless of whether this combination marks singular or plural. Approximately two-thirds of the cases show this association between longer and lower. Once again, instances of this association are roughly equally divided between using it to mark singular and using it to mark plural.
- iv. With regard to Tone, there are a number of very clear sub-regularities. The most noteworthy is that in more than 80% of the words in our sample, at least one of the two forms has L tone.
- v. As can be inferred from Tables 2 and 3, by far the most common combination of tones in number marking is L in one form and H in the other (142 cases, or nearly 40% of our sample). These are about equally divided between nouns in which it is the singular that has L and nouns in which the plural has L; once again, cf. section 1.2.
- vi. When the combination of L and H tones occurs with nouns in which number is also marked by a difference of Length (95 cases), the H tone almost never occurs on the form with the greater degree of Length. That is, in these nouns, L tone and greater Length are closely associated.
- vii. In the same way, in cases where one of the forms has a falling (HL) tone and there is also a difference of Length (24 cases), the HL tone almost always occurs in the form with the greater degree of Length.
- viii. Finally, where there are differences in the coda consonant involving an alternation between /t̪/ and some other consonant (which in our corpus occurs only with monosyllables), the /t̪/ is almost invariably used in the plural form (10 cases out of 11).

This list is not intended as exhaustive, but it is suggestive of the kinds of correlations and associations that can be readily observed and that might help the learner. At the same time, the extent of the irregularity is still noteworthy, and there is no evidence of a default strategy that might be regarded as rule-governed. Rather, the generalizations that we find resemble the kinds of probabilistic sub-regularities within the group of irregulars that have been discussed for English and German past tense formation by e.g. Pinker (1999) and Laaha et al. (2006). Moreover, most of them are generalizations about number marking rather than specifically about plural marking; with the exception of generalization (viii), they are better seen as generalizations about the relation between the marked and unmarked forms rather than between the plural and the singular. The learner must also master the semantic basis of the tripartite number marking system (section 1.2) in order to use these generalizations in acquiring the Dinka lexicon.

## **2.5. Derived nouns and loan words.**

We collected data on number marking for 10 recent loan words and 17 derived nominals of various types. The derived forms are of several types: 4 cattle nouns, 6 deverbal or deadjectival nominalizations, 5 derived forms that might be called agent nouns, and 2 possibly imitative or onomatopoeic nominalizations with reduplicated stems. The loan words are recent borrowings, mostly from Arabic and English. We are well aware that this is an extremely limited sample of morphologically complex and non-native nouns, but it is adequate for our primary purpose, which is to explore the constraints on irregularity rather than to give an exhaustive account of number marking in Dinka.

Cattle nouns are derived from attribute terms that denote the colours or patterns of hides; for bulls, the nominalization involves the prefix /ma-/mi-/ (SG/PL). (It may be relevant to point out that traditional Dinka economy, society and folklore are all heavily based on the keeping of cattle.) The four examples in our corpus are presented in 7.

(7)	Singular	Plural	
a.	ma-kwăc	mi-kwęc	‘spotted bull’
b.	ma-jḵok	mi-jḵok	‘black bull with white chest’
c.	ma-kḵol	mi-kḵol	‘red-brown and white patterned bull’
d.	ma-cḵaar	mi-cḵaar	‘black bull’

As seen from these examples, the marking of number in cattle nouns involves the same mechanisms familiar from the basic nouns, such as differences of Tone, Length, and vowel Height. There are too few cases to determine whether there might be sub-regularities within the set.

The second group of nominalizations are formed from adjective/verb stems by the addition of the prefix /kḵ-/ka-/ (SG/PL), yielding an often abstract noun. The six examples in our corpus are shown in 8.

(8)	Singular	Plural	
	kḵ-pjàṭ	ka-pjàṭ	‘something good’ (< pjàṭ ‘good’)
	kḵ-còol	ka-còl	‘something black’
	kḵ-rěem	ka-rěem	‘pain’
	kḵ-wěṅ	ka-wěṅ	‘something old’
	kḵ-wḵiir	ka-wḵiir	‘water spirit’
	kḵ-ràac	ka-ràc	‘snake’ (< ràc ‘bad’)

The plural formations here are largely regular, though complex, and quite specific to this set of nouns. The prefix has a singular form /kḵ-/ and a plural form that varies between /ka-/ and /ka-/ according to whether the following stem has breathy voice quality or not. The stems themselves show two different patterns of behavior: some are invariable, while others vary in vowel length, normally longer in the singular and shorter in the plural. This pattern of length variation is seen in many free adjective forms when they are in construction with a noun (e.g. /tḵim còol/ ‘black tree’ vs. /tḵiim còl/ ‘black trees’).

The third group of nominalizations consists of agent nouns formed from verbs or other noun stems by the addition of the prefix /mi-/. Our five examples are shown in 9.

(9)	Singular	Plural	
a.	mi-kḵook	mi-kḵook	‘greedy person’
b.	mi-jḵaal	mi-jḵaal	‘father of girls’
c.	mi-kḵoor	mi-kḵoor	‘adulterer’
d.	mi-kwḵiij	mi-kwḵiij	‘miser’
e.	mi-wḵat	mi-wḵaat	‘father of boys’

It can be seen that these all follow generalizations (iv) and (vi) in section 2.4, namely that at least one of the two forms in each pair involves L tone and that the L tone goes with the form

that has the longer vowel if a difference in vowel Length is involved. More specifically, in fact, in four of the five cases we have a plural form with a long (/VVV/) vowel and a L tone. Whether by coincidence or not, this combination reappears in most other reduplicated forms and in nine of the ten loan words. The reduplicated forms in 10 and the loan words exemplified in 11 show that most of the polysyllabic words in these groups have the plural form with a long (/VVV/) vowel and L tone.

(10)	Singular	Plural	Gloss
a.	ma-circìr	ma-circìiir	‘bicycle’
b.	ma-tùktùuk	ma-tùktùuuk	‘car’
(11)	Singular	Plural	Gloss
a.	tùkùul	tùkùuul	‘school (< Eng.)’
b.	galàm	galàaam	‘pen (< Ar. <i>galam-aglaam</i> )’
c.	garnêet	garnèeet	‘grenade (< Eng.)’
d.	kumbâaj	kumbàaaj	‘cup (< Ar. <i>kubbaaya</i> )’
e.	tâaan	tǽen	‘plate (< Ar. <i>ṣaHan-ṣoHoon</i> )’

The only exception to this pattern among the loanwords is the monosyllabic 11e, which has a relatively rare combination of parameter differences typical of the native monosyllables. The other words in 10 and 11 are polysyllabic, and deviate phonologically from the segmental template for native polysyllables. It may be that, as a group, these words, together with the agent nouns, have developed or are developing their own sub-regularity (long (/VVV/) vowel plus L tone in the plural) for number marking. This sub-regularity would have a status in Dinka comparable to the use of the /-s/ plural suffix in German, which has been discussed by various authors. But even here, two out of our 18 cases constitute exceptions to this pattern, and we note that loanwords are traditionally rather rare in Dinka and may still be disfavored.

### 3. Discussion and conclusion.

Our dataset of 400 Dinka nouns yields 82 combinations of phonological differences that can be used to mark the distinction between singular and plural. Some patterns appear with greater frequency than others: that is, patterns involving some combination of differences in Tone, vowel Length, and vowel Height are more common than other combinations. However, though there are clearly some tendencies and probabilistic generalizations about how the phonological differences can be combined, it does not appear possible to identify any phonological or semantic motivation for the choice of number marking pattern for a given noun. Only in some of the derived nominals and loanwords do we see something like a pattern that we might consider to be regular: in the deverbal/deadjectival nouns we see a complex but consistent pattern involving separate singular and plural prefixes, while in agent nouns, reduplicated forms and loanwords we find that most plural forms have a long (/VVV/) vowel and L tone.

Similar degrees of irregularity have been reported for other Western Nilotic languages (e.g. Welmers 1973, Frank 1999, Gilley 2000, Storch 2005). The best-studied case in this context is Nuer, which together with Atuot is the language most closely related to Dinka. Here, as in Dinka, stem-internal changes predominate in the marking of number, but, unlike Dinka, Nuer also has a suffixal marking, which shows up in about 20% of nouns (Frank 1999:9-10). Studies on Nuer noun morphology similar to our own have been reported in Welmers (1973) and Frank (1999). Both report finding several dozen patterns of inflectional marking,

primarily on the basis of stem-internal changes. Frank writes (1999:27): ‘This invokes the possibility that there are two forms (or principal parts) stored for all nouns, a singular and a plural stem from which all forms are derived.’ Consistent with this idea, we note that native speakers are often reluctant to speculate about the plural of unattested nouns (i.e. to participate in ‘wug’ tests; cf. Gleason 1958). They also appear quite willing to acknowledge when they are uncertain about the plural of specific nouns with which they are relatively unfamiliar (e.g., for male speakers, specialized cooking implements).

The psycholinguistic implications of Dinka number marking would certainly be worth exploring. Nothing is known, for example, about the way noun number morphology is acquired by children. Nor is it known whether the results of experimental investigations such as those carried out on speakers of European languages by Bybee and Newman or by Albright would be at all comparable for Dinka speakers. Needless to say, the practical difficulties in pursuing this research would be considerable. Slightly more feasible would be attempts to model the acquisition of Dinka number marking by a connectionist network or other learning algorithm; however, this would be limited by the availability of large amounts of reliable data (cf. our comments on the difficulties of assembling a corpus as relatively small as 400 singular-plural pairs). In any case there is a rich potential for future research.

Finally, it is worth noting that, both in Dinka and in other Nilo-Saharan languages, irregularity is endemic in the noun morphology, but not in the verb morphology (cf. Andersen 1993, Frank 1999). While the verb morphology typically deploys the same structural and phonological resources (viz., in the case of Dinka, segmental and especially suprasegmental stem changes), it is essentially regular once a limited set of conjugation classes is identified. Systematic investigation across languages is needed to determine whether or not this is more than an accident due to internal development within the Nilo-Saharan language family. The data reported here, like reports on other Nilo-Saharan languages, clearly suggest that ‘extremes in irregularity ... are not the cognitive burden that we may have wanted to postulate for theoretical linguistic reasons’ (Frank 1999:68). At the same time, though, the burden of storing a pair of forms for every noun is relatively light compared to what would be involved in learning idiosyncratic forms for every form of every verb in a paradigm like the one in Table 1. That is, it seems reasonable to speculate that, cross-linguistically, there may be more functional pressure towards regularity in verbs – which typically have many forms in a highly inflected language – than in nouns – which may be inflected only for a few categories such as number, as in Dinka. This might be taken as evidence in favor of some sort of dual-route view, or at the very least for Pinker's contention (1999: 18) that rule-based processing is a cognitive necessity in complex inflectional systems. In effect, Dinka noun number inflection certainly proves that the limits of morphological irregularity lie considerably beyond what we find in German or Hebrew, but it does not necessarily show that regularity can be dispensed with altogether.

### *References*

- Albright, Adam. 2002. Islands of reliability for regular morphology: Evidence from Italian. *Language* 78, 684-709.
- Andersen, Torben. 1990. Vowel length in Western Nilotic languages. *Acta Linguistica Hafniensia* 22, 5-26.
- Andersen, Torben. 1993. Vowel quality alternation in Dinka verb inflection. *Phonology* 10, 1-42.

- Andersen, Torben. 2002. Case inflection and nominal head marking in Dinka. *Journal of African Languages and Linguistics* 23: 1-30.
- Baayen, R. Harald. 2007. Storage and computation in the mental lexicon. In G. Jarema and G. Libben (eds.). *The Mental Lexicon: Core Perspectives*. Elsevier, 81-104.
- Bybee, Joan L. and Jean E. Newman. 1995. Are stem changes as natural as affixes? *Linguistics* 33: 633-654.
- Dimmendaal, Gerrit J. 2000. Number Marking and Noun Categorization in Nilo-Saharan Languages. *Anthropological Linguistics* 42: 214-261.
- Frank, Wright Jay. 1999. Nuer noun morphology. MA dissertation, State University of New York at Buffalo.
- Gilley, Leoma G. 2000. Singulars and Plurals in Shilluk – a search for order. *Occasional Papers in the Study of Sudanese Languages* 8.
- Gleason, Jean Berko. 1958. The child's learning of English morphology. *Word* 14:150-77.
- Laaha, Sabine, Dorit Ravid, Katharina Korecky-Kröll, Gregor Laaha and Wolfgang U. Dressler. 2006. Early noun plurals in German: regularity, productivity or default? *Journal of Child Language* 33: 271-302.
- McClelland, James, and Karalyn Patterson. 2002. Words or rules cannot exploit the regularity in exceptions: reply to Pinker and Ullman. *Trends in Cognitive Sciences* 6: 465-472.
- Pinker, Steven. 1999. *Words and Rules: The Ingredients of Language*. New York: Basic Books.
- Pinker, Steven, and Michael Ullman. 2002. The past and future of the past tense. *Trends in Cognitive Sciences* 6: 456-463.
- Ravid, Dorit and Rachel Schiff. 2009. Morphophonological categories of noun plurals in Hebrew: a developmental study. *Linguistics* 47: 45-63.
- Reh, Mechthild. 1999. *Anywa-English and English-Anywa Dictionary*. Cologne: Rüdiger Köppe Verlag.
- Remijsen, Bert and Caguor Adong Manyang. 2009. Luanyjang Dinka. *Journal of the International Phonetic Association* 39: 113-124.
- Remijsen, Bert and Leoma G. Gilley. 2008. Why are three-level vowel length systems rare? Insights from Dinka (Luanyjang dialect). *Journal of Phonetics* 36: 318-344.
- Remijsen, Bert and D. Robert Ladd. 2008. The tone system of the Luanyjang dialect of Dinka. *Journal of African Languages and Linguistics* 29: 173-213.
- Rumelhart, David, and James McClelland. 1986. On learning the past tense of English verbs. In J. McClelland, D. Rumelhart, and the PDP Research Group (eds.), *Parallel Distributed Processing*, vol. 2, pp. 216-271. Cambridge MA: MIT Press.
- Storch, Anne. 2005. *The Noun Morphology of Western Nilotic*. Nilo-Saharan vol. 21. Cologne: Rüdiger Köppe Verlag.
- Welmers, William. 1973. *African Language Structures*. Berkeley: University of California Press.