Number in Dinka and ir(regularity)

Ladd et al. 2009

0 Introduction

- Morphological processes are routinely classified as regular or irregular, so a bit of conceptual analysis is in order.
- The goal: figure out what linguists mean by regularity *independently of its status in the natural world*. Does the intuition break down somewhere? If so, this would definitely be relevant for theories that make reference to the notion.
- Call the set of environments in which a process M applies the "morphological class for M."
- Two properties seem to matter when considering a process M: i) the size of its class (which ultimately determines the type frequency of M), and ii) the naturalness of its class.
- In other words: i) do we want a generalization? and ii) how easy/hard is it to state this generalization, given that we want it?
- These two properties are not independent: a class that includes a single environment (the kind for which we don't want a generalization) is always maximally natural.
- Let's test our intuitions:

(1) Morpheme with alternants M_1, M_2

Alternant	Class	Naturalness	Size	Type frequency
M_1	Animates	High	$\|A\|$	0.5
M_2	Inanimates	High	$\|\overline{A}\|$	0.5

As long as its class is natural, we are fine with calling a process regular.

- We usually call a process irregular when its class's membership conditions are no longer natural.
- The exception to this is the default/elsewhere process, which is associated with the largest unnatural class, M_1 below.

(2)

Alternant	Class	Naturalness	Size	Type frequency
M_1	80% of animates +	Low	$0.8\|A\|+$	0.7
	60% of inanimates		$0.6\|\overline{A}\|$	
M_2	20% of animates +	Low	$0.2\|A\|+$	0.3
	40% of inanimates		$0.4\|\overline{A}\|+$	

- So:
  - Process is regular if it targets a natural class.
  - Process is regular if it targets the largest unnatural class.
- What happens to our intuitions when every unnatural class is roughly of the same size?

(3)

Alternant	Class	Naturalness	Size	Type frequency
M_1	60% of animates +	Low	$0.6\|A\|+$	0.5
	40% of inanimates		$0.4\|\overline{A}\|$	
M_2	40% of animates +	Low	$0.4\|A\|+$	0.5
	60% of inanimates		$0.6\|\overline{A}\|+$	

There is no basis for picking a default, and we are hard-pressed to call either of M_1 and M_2 regular.

- Naturalness depends on some kind of evaluation metric. (With enough primitives and operators, it's possible to define any class.)
- Usual distinction between regularity and subregularity not principled: calling a process subregular simply means finding membership criteria for a seemingly unnatural class.
- Is our use of terms like regular and irregular betraying expectations we have have accumulated through exposure to particular languages/phenomena?
- This is important for theories of morphological processing that claim that only regular processes are productive, or that only regular processes are *processes* properly speaking (the dual-route model, Pinker 1999, Pinker & Ullman 2002).

1 Number marking in Dinka

- Ladd et al. present an overview of the Dinka (Nilo-Saharan) number marking system. In their 400-word sample, they find 82 distinct phonological alternations in use to mark the plural-singular distinction.
- By the standards just discussed, almost all of these alternations would be irregular, because of i) their low overall frequency, and ii) the opacity of their environment.
- Singular/plural pairs for 400 words were coded for the parameters in (4).

(4)

PARAMETER	DEGREES OR TYPES
Vowel length	3 (same, different by 1 degree, different by 2 degrees)[7]
Tone	7 (same, L ~ H, L ~ HL, L ~ LH, H ~ HL, H ~ LH, HL ~ LH)
Coda	3 (same, C ~ t̪, other difference)
Vowel height	2 (same, different)
Vowel backness	2 (same, different)
Breaking	2 (same, different)
Voice quality	2 (same, different)

- One important methodological point is that the direction of the change was not coded. Thus a tone (L ~ H) alternation applies to both cases where it is the plural that's H, and cases where it is the singular.
- Had directionality been coded, the number of alternations would be even greater.
- Nilo-Saharan languages have what is called a tripartite number marking system, which relates to the mass-count distinction.
- For nouns that usually come in masses or groups, the singular is the marked form. For some other nouns, the plural is marked, and there is yet another category of nouns that are marked for both plural and singular.
- Assuming speakers can make this semantic distinction, *all* they then need to learn is how to mark "number."
- Reasonable coding for single parameters, then.
- But for combinations, potentially valuable information gets swallowed up. For instance, the combination length (1) + tone (L ~ H) has the four subcases we get from crossing the directionality value of each parameter.
- Is every single one of these four cases attested? Hopefully not!

2 Results

- Breakdown of the 400 pairs: i) 312 native monosyllabic words, ii) 61 native polysyllabic (monomorphemic) words, iii) 17 nominalizations and morphologically complex words, iv) 10 recent loan words.
- **Native monosyllabic words**: 81 distinct combinations of parameters.
- 85% involve a tone difference, 71% a length difference, 49% vowel height, 21% vowel breaking, 11% voice quality, 11% coda consonant, 1% vowel backness.

(5)

PATTERN	%	*N*
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

- **Native polysyllabic words**: 33 combinations, all but one shared with monosyllabic words.

(6)

PATTERN	%	*N*
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

- **Agent nouns derived from verbs, and new loanwords** show some kind of subregular pattern: long VVV vowel and low tone in the plural.

(7)

	SINGULAR	PLURAL	
a.	mi-kŏok	mi-kòook	'greedy person'
b.	mi-ɲàal	mi-ɲàaal	'father of girls'
c.	mi-kɔ̀ɔɔr	mi-kɔ̌ɔɔr	'adulterer'
d.	mi-kwǐiiŋ	mi-kwìiiŋ	'miser'
e.	mi-wà̰t	mi-wà̰aat	'father of boys'

(8)

	SINGULAR	PLURAL	GLOSS
a.	t̪u̪kû̪ul	t̪u̪kù̪uul	'school' (< Eng.)
b.	galàm	galàaam	'pen' (< Ar. *galam-aglaam*)
c.	garnêet	garnèeet	'grenade' (< Eng.)
d.	kumbâaj	kumbàaaj	'cup' (< Ar. *kubbaaya*)
e.	t̪â̪aan	t̪ɛ̪̌ɛn	'plate' (< Ar. *ṣaHan-ṣoHoon*)

- Because directionality was not coded, it's hard to recover correlations in directionality.
- But the authors offer some comments. For example, in their discussion about native monomorphemic words, "when the combination of L and H tones occurs with nouns in which number is also marked by a difference of length (ninety-five cases), [...] **L tone and greater length are closely associated.**"
- Although none of the singulars in (7) and (8) involve a H tone, they do match the bolded observation.

3 Discussion

- Lots of moving parts. No doubt some of these processes could be collapsed if...
  - some interactions were attributed to phonology.
  - different coding conventions were adopted.
- In particular, it's clear that parameters with a lot of levels like tone (7) are responsible for the proliferation of possible combinations.
- We may get a more informative picture by not treating these variables categorically.
- The data still present a challenge for intuitions about the concept of (ir)regularity, and consequently for any theoretical stance that takes it for granted.