UCLA Working Papers in Linguistics, vol.0, June 2000
Papers in Lakhota—Munro (ed.)

THE PRODUCTIVITY OF INFIXATION IN LAKHOTA[*]

ADAM ALBRIGHT
aalbrigh@ucla.edu

A growing body of work on morphological irregularity has shown that the productivity of irregular morphological processes is sensitive to phonological context. However, these studies have focused almost exclusively on patterns of irregularity found in European languages. In this paper, I discuss a typologically different pattern of irregularity found in Lakhota: person marking appears as a prefix for some verbs, and as an infix for others. I combine computational modeling of the Lakhota lexicon and an experimental "wug" test, to show that infixed person marking is highly sensitive to context, just like other types of morphological irregularity.

1. INTRODUCTION

It is not uncommon in the world's languages for a particular morpheme to surface variably as an infix or as a prefix/suffix. One of the insights of McCarthy and Prince (1993), further developed in the OT literature, is that the decision to infix is often driven by prosodic considerations in the phonology. For example, possessive markers in Ulwa are placed after the first foot of the root, with the result that they are infixed in roots that are longer than one foot, and suffixed in words which contain just one foot. A second consideration which can drive infixation is syllable structure. An example of this is the morphemes *-um-* and *-in-* in Tagalog, which are prefixed to vowel-initial roots, but infixed in consonant-initial roots to create CV syllables and avoid VC syllables. One final consideration which has been argued to drive variable infixation is the local segmental environment. For example, Crowhurst (1998) points out that in at least some varieties of Toba Batak, the *-um-* morpheme which is cognate to Tagalog *-um-* is infixed only after certain consonants (/t k s d dʒ g r l h/).

In all of these cases, the analytical goal has been to make sense of a predictable distribution (infix vs. affix) in terms of its satisfaction of prosodic phonological constraints. In this paper, I will consider a case which has received less attention in the literature: morphemes which

[*] I would like to thank our Lakhota teacher Mary Iron Teeth, who patiently provided difficult judgments about silly made-up words and about unusual uses of actual words — her good-natured willingness to play the "elderly aunt" game week after week has been truly remarkable. I would also like to thank Pam Munro, whose generous willingness to share her Lakhota Verb List made this analysis possible. Comments and suggestions are gratefully welcomed.

may surface *unpredictably* as prefixes or infixes. The example I will focus on is person marking in Lakhota, as illustrated in (1):

(1) Variable position of first singular subject marker *-wa-*:

a. Prefixed:

lówan	'he sings'	~	*wa-lówan*	'I sing'
núwe	'he swims'	~	*wa-núwe*	'I swim'
káge	'he does/makes'	~	*wa-káge*	'I do/make'

b. Infixed:

máni	'he walks'	~	*ma-wá-ni*	'I walk'
aphé	'he hits'	~	*a-wá-phe*	'I hit'
hoxpé	'he coughs'	~	*ho-wá-xpe*	'I cough'

Such morphological irregularity poses not only an analytical problem for the linguist, but also a learning problem for the language learner. At the most basic level, for each word the learner must learn whether that word takes prefixation or infixation. However, this task might be simplified considerably if one could predict the behavior of words somehow, and if one pattern could be designated as a "default" to cover the majority of cases. In order to do this, we must be able to generalize over individual lexical items in some fashion. One scheme for how children might do this is the "Minimal Generality" hypothesis, outlined out by Albright and Hayes (1998) — namely, that children attempt to make sense of apparently arbitrary morphological patterns by paying careful attention to the phonological form of the roots involved. Under this view, the acquisition process is seen as a deliberate search for patterns which help predict class membership; as Zubin and Köpcke (1981, p.439) phrase it, "[w]e know that children are not passive consumers of morphological irregularity." The end result is that languages have phonological regularities which line up with morphological irregularity, to produce a class system where membership is sort of — but not completely — arbitrary.

The relationship between morphological irregularity and phonological form has been investigated for languages like English (Prasada and Pinker 1993) and for Italian verb classes (Albright 1998), but it has never been tested in a language with variably positioned morphemes. The goal of this paper, therefore, is to extend this line of investigation to Lakhota, in order to test the influence of phonology on morphology in a typologically different language. I will start with an overview of the place of affixation in existing Lakhota words, offering evidence that it is in fact a form of morphological irregularity, and reviewing some insights from Boas and Deloria (1941) on this matter. Next, I will outline an attempt to discover statistical tendencies computationally, using an automated pattern-discoverer developed by Albright and Hayes (1998). The goal of this section will be to uncover “islands of reliability” — that is, phonological neighborhoods in which words are especially likely to prefix or infix. I will then try to show that native speakers are in fact aware of these “islands,” and that phonological form does play a major role in shaping intuitions about how novel words should inflect.

2. The Position of Subject Markers in Lakhota

2.1. *Overview of subject markers*

Subject agreement is marked in Lakhota by person and number affixes; the person affixes are either prefixed or infixed, while the plural marker *-p* is suffixed. The person markers can be divided into two series, which I will label as “I” and “II” (following Munro 1989).[1] The subject markers for all persons are shown in Table 1.

Table 1: The subject marking affixes of Lakhota

“Series I”				“Series II”	
		(*y*-initial stems)			
wa	*un(k)*[*]*...p*	*bl*	*uny...p*	*ma*	*un(k)*[*]*...p*
ya	*ya...p*	*l*	*l...p*	*ni*	*ni...p*
∅	*∅...p*	*y*	*y...p*	∅	*∅...p*

[*] 1pl *-un-* includes *k* before a vowel

Boas and Deloria (1941) state that “[i]t is not possible to give absolutely consistent rules for the position of personal pronouns” (i.e., affixes) (p.78). As seen from example (1) above, the marker is prefixed for some verbs, and infixed for others. Therefore, the position of subject marking is to a certain extent unpredictable, and must be listed on a verb-by-verb basis.

It is natural at this point to wonder how we can be so sure that the position of subject marking is not driven by phonotactic pressures in Lakhota. After all, the verbs in (1a) and (1b) are similar in structure, but they are not exactly minimal pairs. Is it possible that the alternation between prefixiation and infixation is in fact conditioned by a phonological environment, and we have simply failed to recognize the correct generalization? Without being able offer exact minimal pairs, I can not rule this possibility out completely; however, I can offer several types of evidence which make it seem unlikely.

[1] The difference between Series I and Series II subject marking can generally be interpreted as an active/stative distinction; however, Munro has pointed out that some Series II verbs are not obviously stative, so the behavior of a verb is not completely predictable from its subcategorization properties.

The first fact which makes a phonotactic explanation unlikely is that verbs with fully parallel structure (stress, manner of articulation, etc.) can take different patterns. Thus, the verbs *lówan* ‘sing’ and *máni* ‘walk’ in (1) are both composed of sonorants and have the same stress pattern, but one takes infixation and the other prefixation. In addition, Lakhota possesses other affixes which have similar phonological structure, but which can never infix. For example, the indefinitive object (valency-reducing) affix *wa-* is homophonous with the Series I first person marker, but it is prefixed for all verbs (including those which take infixed subject marking). Therefore, we can infer that it would be phonotactically legal to prefix the subject markers for all verbs.

Finally, it appears that for a certain number of verbs, the subject markers may even be allowed in more than one position. In (2) we see that the verb *washícu* ‘be a white man’ allows either prefixation or infixation, while the verb *wichásha* ‘be a man’ allows only infixation.

(2) Variable affix positions for some verbs

a.	*washícu*	‘be a white man’	
		wamáshicu	‘I am a white man’
		mawáshicu	‘I am a white man’
b.	*wichásha*	‘be a man’	
		wimáchasha	‘I am a man’
		**mawíchasha*	‘I am a man’

The two verbs in (2) are quite similar phonologically, but they have different morphological properties — therefore, it really does seem that the position of subject marking can not be reduced to a phonotactic effect.

A second hypothesis which should be considered is a morphosyntactic one, in which infixing verbs are considered complex somehow. One approach along these lines would be to analyze infixing verbs as compounds, and then ensure that subject marking is prefixed to the second verb (the head of the compound) before the first verb is compounded to it. This approach would run into problems, however, with affixes that are always prefixed, such as the indefinite object *wa-*. We would have to say that these affixes are added *after* the compound is formed, but these suffixes typically affect subcategorization requirements (usually by changing valency), and subcategorization requirements are a property of the *head* of the compound — so why should they be added to the complement? In addition, inflectional elements like person marking are typically the last elements to be added in a verbal complex, so it would be quite surprising to find a language where person marking precedes valency-changing derivational morphology. A second approach would be to say that infixing verbs are composed of a prefix plus a verb, and allow inflectional material to be inserted between the prefix and the verb — much the same as separably prefixed verbs in German or Dutch. In this case, infixation would be reduced to a problem of morpheme-ordering (valency morphemes, “separable prefix” parts of the verb, subject marking, and finally the verb itself).

Boas and Deloria do in fact suggest that a large number of infixing verbs, which begin with the vowels *a*-, *i*-, and *o*-, can be analyzed in this way (p.79). They propose that all initial *a*-, *i*-, and *o*- must be locative prefixes, which may indeed be true etymologically. However, not all of these prefixes are productive synchronically, and their distribution and semantics are not at all obvious. In particular, many verbs with initial *a*-, *i*-, and *o*- do not possess counterparts without the initial vowel, and they do not contain any apparent locative meaning. Therefore, it is not clear how speakers would learn what is essentially a historical fact about their language. Furthermore, not all infixing verbs begin with one of these etymological prefixes, so even if we accept the prefix analysis for initial *a*-, *i*-, and *o*-, we would still need to explain infixation in other verbs.

We can see then that infixation in Lakhota does not seem to have a straightforward morphosyntactic analysis, nor is there an obvious phonological motivation. I conclude, therefore, that it is best treated as a form of morphological irregularity.

This is not to say, however, that there is no relation between phonological form and infixation. Despite the basic unpredictability of the position of subject marking, Boas and Deloria do make a number of observations about which verbs seem most likely to prefix or infix. The first observation that they make is that monosyllabic verbs always prefix — i.e., a syllable is never broken up by the insertion of an infix (p.78). They go on to observe that all polysyllabic verbs which begin with a vowel infix the marker immediately after the initial vowel, as shown in (3). (They do list two exceptions, however, so in fact this is only true in most cases.)

(3) V-initial polysyllabic roots (3sg → 1sg)

a. Subject marking infixed after V

ixá	*iwáxa*	‘smile’
ómna	*ówamna*	‘smell’
áphe	*awáphe*	‘hit’

b. Exceptions (prefixed)

únpa	*wa’únpa*	‘lie down’
óta	*waóta*	‘be many’

As mentioned above, Boas and Deloria claim that these initial vowels can probably all be analyzed as the locative prefixes ‘*a*-, ‘*i*-, and ‘*o*- (p.79), although there is not strong evidence for this in all cases.

An observation which is not so easily reduced to morphosyntax is the generalization that many verbs beginning with *th*- infix the subject marking (§83), as in (4):

(4) Initial *th*-V- encourages infixation

a. Infixed after th-V-

*tha*phá*	‘follow’ (I)
*tha*kpé*	‘attack’ (I)
*tha*ó*	‘wound by shooting’ (I)
*tho*kshú*	‘haul away’
*the*mní*	‘sweat’ (II)

b. BUT: some th-V- exceptions (Buechel 1970)

thamáhecha		*mathámahecha*	‘lean’
thanín	~	*mathánin*	‘be visible’
thánka	~	*mathánka*	‘big’
(etc.)			

Boas and Deloria provide quite a large number of such phonological generalizations, including that verbs beginning with *m*- are mixed (some infix and some prefix, §83), verbs ending with -*pha* all belong to Series I and infix right before the -*pha* (§90), and verbs ending with -*kha* are all prefixing (§91).[2]

It is significant that Boas and Deloria considered it worthwhile to devote so much attention to the phonological environments surrounding infixation. In fact, the environments which they focus on are those which they felt to be the “islands of reliability” for various patterns, and these intuitions must have been supported by careful consideration of a large number of words and their place of affixation. This gives us nice *a priori* evidence that phonological form may play an important role in determining the place of affixation, and that we may be able to observe its effects in decisions about novel words.

It is also worth noting that Boas and Deloria do *not* designate any particular pattern (prefixation or infixation) as a “default” pattern. This may be due in part to the fact that Lakhota has few loanwords, so one of the most common “default” contexts does not exist in Lakhota. In fact it is probably *not* the case that we want to call prefixation the

[2] Boas and Deloria speculate that the final -*pha* must actually be a frozen verb of some sort, but they can not find any coherent meaning for it which would unite all -*pha* final verbs. Therefore, although this speculation may be true historically, there does not seem to be any synchronic evidence for it.

default, since speakers are completely comfortable with prefixation in some novel words (e.g. *mathó* 'I am blue') and infixation in others (e.g., *i-ma-camna* 'I am snowing').

2.2. *A Method for Locating "Islands of Reliability"*

In order to test for islands of reliability such as those uncovered by Boas and Deloria, I submitted a database of existing Lakhota verbs to the Automated Learner algorithm developed by Albright and Hayes (1998). This algorithm, which was designed in part to learn morphological generalizations in the face of exceptions and competing patterns, goes through the data set and collects statistics about the phonological neighborhoods in which each morphological pattern applies. For example, two verbs which infix are *máni* 'walk' and *mánu* 'steal,' so the algorithm would compare these words to discover that infixation is possible after *ma-* and before *n* plus a high vowel:

(5) Comparison of *máni* 'walk' and *mánu* 'steal'

a.	*comparing:*	→ [wa] /	ma	__	n	i
b.	*with:*	→ [wa] /	ma	__	n	u
c.	*yields:*	→ [wa] /	ma	__	n	é *+syl* ù ê*-cons* ú ê+*high*ú ëetc... û

The program proceeds by comparing *all* of the words of the language, considering what material is shared by each pair of verbs and locating the phonological neighborhoods of the language. Furthermore, these neighborhoods have *reliability statistics* attached to them, in the form of statements like "five out of seven of the existing words which contain [man] + a high vowel take infixed subject markers between [ma] and [n]." According to the Albright and Hayes model, these reliability statistics, which serve as a batting average for generalizations, are used by language learners to locate the islands of reliability for the morphological patterns of their language, and they provide the basis for well-formedness intuitions about novel words.

The infixation environments of Lakhota were tested by first creating a database of 824 Lakhota verbs, in the third singular (no subject marker) and the first singular (-*wa*- or -*ma*- subject marker). The majority of these verbs were taken from Munro (1989). I included both "base" and "derived" forms (e.g., verbs like *ixa* 'laugh' and *a'ixa* 'laugh at' were both included), and I also included verbs with anomalous inflections (such as *mánke* 'I sit' from *yánke* 'he sits'). I did exclude verbs with multiple changes (such as *iblable* 'I go' from *iyaye* 'he goes'), since the morpheme parser in the Albright and Hayes learner is unable to parse multiple change locations. I also excluded verbs whose exact affixation position was not clear from Munro (1989), and I excluded verbs which I knew that Mary Iron Teeth did not use. (I did not confirm the entire list, however, since in large part her judgments matched those in the list.) The list was also augmented to include a more representative sample of verbs in some environments of interest, such as those beginning with *o*- and *i*-, and those beginning with *th*-.

This database was then fed to the Albright and Hayes learner program, resulting in a comprehensive list of all of the phonological neighborhoods in the input file, along with the likelihood to prefix or infix in those environments. The next step was to use this list to find the "islands of reliability," or environments which are especially likely to prefix or infix. This was accomplished by creating a list of di- and tri-syllabic nonsense words, by randomly combining different consonants (either alone or in phonotactically legal clusters) and vowels to create 2,493 nonsense words, as in (6).

(6) Sample of nonsense test words

kake
khake
chake
shake
...

ptaxaye
slaxaye
maxaye
glaxaye
...

iyokagle
iyokhagle
iyochagle
...

palapte
paglapte
payapte
...

For each of these words, the computer outputted a variety of guesses, including prefixation and infixation in different positions. In addition, it gave a numerical score for each output, reflecting the program's "confidence" in that guess given the patterns in the input file. These outputs were then imported into Excel and sorted by confidence, in order to find the novel words which should almost certainly be prefixing or infixing.

2.3. *Results of automated analysis*

The "islands of reliability" located by this procedure were generally *not* the same ones which Boas and Deloria isolated in their discussion of the problem. (In fact, it appears that this is often the case — careful linguists may notice a small number of beautiful salient generalizations, while the program's comprehensive search procedure allows it to find hundreds of other arcane generalizations which are statistically just as reliable in the lexicon.)

The most trustable pattern in the entire input database was that *y*-initial verbs changed to *bl*- when the following vowel was a *u*-, as in (7). (The '...' indicates that any amount of segmental material can appear to the right of the initial *yu*.) This generalization worked almost 90% of the time, making it a rather good island of reliability.

(7) The best island: $\boxed{\text{y} \rightarrow \text{bl / \#\_\_u…}}$ .858
(batting average: 22/24, including *yuha*, *yu'ile*, *yu'onihan*, *yu'ota*, *yugho*, *yuha*, *yuja*, *yushka*, *yushla*, *yushna*, *yushpi*, etc…, missing only *yush'inyen*, *yush'inyeye*)

In fact *yu-* is an instrumental prefix in Lakhota, so it is not an accident that there are so many *yu-* initial verbs and they all pattern together. Note that this is probably *not* the absolutely most reliable pattern in the entire language, because the database for this simulation was rather small and verbs beginning with *yu-* were probably overrepresented. However, a glance through Buechel (1970) does reveal that this is a relatively solid generalization about Lakhota verbs.

One island of reliability for infixation is after *o-* and before *-gla*, as in (8). This is a subcase of Boas and Deloria's generalization that vowel-initial words infix, and this is clearly due to the fact that *o-* may function as a productive prefix. However, as it turns out, this infixation is *especially* likely when the following material is *gla*.

(8) An island of reliability for infixation:
$\boxed{\quad \rightarrow \text{wa / …o\_\_gla…}}$ .852
(supported by: *oglake*, *woglake*, *iwoglake*, *oglaxnigha*, *oglapshun*, *oglapta*)

We are also able to locate reliable environments for prefixation, such as before *pa-* as in (9). In fact, *pa-* is an instrumental prefix which should always come after the subject marking (Boas and Deloria, p.45), so we see that in a sense these islands are modeling facts about affix ordering by brute force phonological generalizations. (I will return to this issue later.)

(9) A good island for prefixation:
$\boxed{\quad \rightarrow \text{wa / \#\_\_pa…}}$ .823
(supported by: *pabla*, *pahi*, *pajaja*, *pajo*, *pakhinta*, *pakize*, *pakse*, *paphope*, *papsun*, *pasi*, *pasise*, *pasleca*, *patitan*, *pawiyakpa*, *paxpe*, *pazo*, *pabu*)

By comparing all environments, the program also comes up with many rather complicated generalizations, such as "infix after *na* when it is preceded by a vowel, and before an obstruent," as in (10). In fact, we can see that the reason this is reliable is a combination of morphology (the vowels and *na-* are prefixes), and phonology (the fact that this is especially reliable before obstruents). So we see again that these crude generalizations are actually capturing a lot of what we might have considered morphosyntactic by brute force in the phonology:

(10) A more complicated island:
→ wa / Vna__ C [-*son*] … .678
("infix *wa* when there is a *na* preceded by a vowel, and before an obstruent")
(supported by: *anapte, inaji, inaxme, inaxni, anaslate, onatha, onaxleca, onaxtake, onajin, onaphe, onashloka, inapa, inapsaka*)

In addition to these very reliable environments, the program is also able to tell us what some especially *unreliable* environments are for particular patterns. For example, *y* does not change to *bl* in any arbitrary place between two vowels, and the generalization which would do this has a very low batting average:

(11) A bad generalization:
y → bl / V__V .122
("change *y* to *bl* between any two vowels")

- works for: *ayushtan, ayuta, eyuthe, iyotake, iyucan, iyukcan, iyuskepe, iyuthe, iyuweghe, oyake, oyuspe, ayuta, oyuze*
- but NOT for: *iya, iyakiphe, iyanke, iyaphe, iyayakhiye, iyaye, iyayeye, iyekichiska, iyeska, iyethokca, iyokihi, iyokphi, iyokphiye, iyophekhiye, iyophey*e + 20 others

The generalization in (11) performs poorly because it tries to extend a pattern into too general an environment — that is, just because a change happens often word-initially (and in particular, before *u*) does not mean that it is generally true before vowels anywhere in the word. A different type of unreliability can arise when a pattern almost never applies in a particular environment, such as prefixing *wa*- before *ik*:

(12) An "Island of UNreliability":
→ wa / #__ik… .123

- works for: *ikpaxpe*
- but NOT for: *ikahi, ikikcu, ikishtece, ikix'an, ikpasise, ikpazo, ikputhake, ikpazo, ikamna*

It would be impossible to discuss all of the islands which the automated analysis considered, since of course there are thousands of possible generalizations which could be made about any language, and this is especially true for infixing languages (since the infix is flanked by phonological environments both on the left and on the right). The important result here is not the individual reliability of various environments, but rather the fact that we can use these reliabilities to try out thousands of nonsense words and locate those which are especially likely to sound good or bad with a particular affix location.

3. THE “WUG” TEST

3.1. *Berko’s wug test*

Once the islands of reliability for prefixation and infixation were located, as described in section 2.2, the next step was to see to what extent this information mirrored the actual knowledge of a native speaker. One common way to test what generalizations speakers have made about their language is the “wug” test, pioneered by Berko (1958). The basic paradigm for the wug test is to present speakers with a novel word in a frame sentence, and then create a situation in which the speaker must inflect the novel word in order to complete a sentence. For example, Berko presented children with pictures of fictional animals in the singular, and required the children to refer to them in the plural by adding the plural suffix -*s*, as in (13).

(13) Berko’s “wug” test
 a. present a novel word in a plausible frame sentence
 “This is a wug”

 b. task requires consultant to inflect “wug” word
 “*Here is another wug. Now there are two _________.*”

 (target: *“wugs”*)

There are many different ways to elicit the inflected novel word, and each technique has its own advantages and drawbacks. I experimented with several different ways of doing this for Lakhota; first, I tried a Berko-like task by simply presenting a sentence with a verb in the 3rd singular (i.e., with no subject marking), and requesting the consultant to fill in the blank in a sentence with a first singular context. This open-response task is difficult for many reasons, however. First, it requires the consultant to “learn” the novel word with very little input — basically after hearing it just once. Also, it seems that for most people, the mechanism for “brainstorming” possible morphological outcomes is rather slow,[3] so the open-response fill-in-the-blank format can be frustrating for consultants. Therefore, I settled on a slightly more elaborate version of this test to use for Lakhota.

[3] I myself have observed this with English and German speakers when requested to brainstorm possible past tenses for a novel verb, and Bruce Derwing (p.c.) points out that when presented with a novel affix, even linguistics grad students have a hard time applying it to novel forms, but when they are presented with possible choices, they are instantly able to choose the one which sounds “correct.”

3.2. *The elderly aunt tells a story*

In order to overcome some of the problems discussed above, I designed a more constrained version of the fill-in-the-blank wug test. The scenario was as follows: you are watching TV with your elderly aunt. There is a documentary on, and you can see in the distance that a man is doing something that you can not quite make out. Your aunt, however, recognizing what he is doing, gets excited and tells you a story about it:

(14) Example: novel verb *okácho* (task is to fill in the blanks)
a. *Héchiya* *he* okácho.
over there he *okácho*-3sg
'That guy *okácho*-s.'
b. *Kaká* ______ *únspe-ma-khiye.*
grandfather *-1sg.OBJ-taught
'My grandfather taught me how to ________ stuff.'
(target: *wakácho, okácho, wa'ókacho*)
INDEF-*okácho* (not the same *wa*- as the 1sg!)
c. *Lehán* *tuwénni* ___________.
nowadays nobody
'People don't __________ *anymore.'*
(target: *okácho-shni*)
okácho-NEG
d. *Miyé nahánxchi* _____.
I sometimes
'Sometimes I still ____*.'*

æ targets:	*wa'ókacho*	*ma'ókacho*	ö
ç	*owákacho*	*omákacho*	÷
ç	*okáwacho*	*okámacho*	÷
è	...	...	ø

The first important feature of this paradigm is that involves several blanks which require either exactly the same form as the inital form in (14), or else a form which has been modified in a completely predictable way (such as adding *-shni* to form a negative in (14)). This gives the consultant a chance to become familiar with the word, and to say it a few times. It also gave me a chance to make sure that the consultant had apprehended the word accurately, so I could repeat the prompt sentence in (a) if necessary.

The second way in which this task differed from Berko's wug task was that I myself presented the possible outcomes for the inflected form (in this case the form in (14), rather than requiring the consultant to make them up by brainstorming. The number of plausible outcomes for each novel verb is large in Lakhota, but it is finite, so I was able to make up a list of *all* possible outcomes and have the consultant rate each one on a scale of 1 to 10. I also found that it was helpful for the consultant to be able to look at all of the outputs on paper while rating them, rather than trying to keep all of them in mind while comparing them.

The result of this task was a list of ratings for each of the possible outcomes for a novel word, as in (15).

(15) Ratings for possible 1sg forms for the novel verb *okácho*

waókacho	5	*maókacho*	0
owákacho	9	*omákacho*	4
okáwacho	0	*okámacho*	0

4. RESULTS

4.1. *Overall properties of the ratings*

Because of the difficulty of treating made-up words as if they were real words, it is always important to ask whether a wug-test has given interpretable results. The first point, and one which I feel should not be overlooked, is that the task outlined in section 3.2 was doable, and in fact it did not seem to be especially frustrating or confusing. (This is also thanks in large part to Mary Iron Teeth's remarkable patience in considering silly word after silly word, and her willingness to express opinions about things which she has never heard — it is certainly not the case that all speakers would be as proficient at this task as she was.) In order to test for session-to-session consistency, I repeated some words after an interval of two or three weeks, and found that there was considerable fluctuation in the judgments. However, this is not at all surprising, since people's judgments about real words and sentences can vary considerably from moment to moment, and as far as I know, no one has ever attempted to study how stable wug-test judgments are from session to session in any language. Therefore I will point this out as a possible problem, but not one which is unique to this particular set of judgments.[4]

Therefore the first question we can ask about the ratings is whether they do in fact match the predictions. When we consider individual test items, the answer here seems to be "sort of, to some extent," as in (16).

[4] One possible response to massive variation might be to see if the ratings are more consistent when considered as relative rankings rather than as absolute ratings from 1-10. I have not pursued this, however, because in fact it did not seem that the ratings would be any more consistent when recoded as rankings rather than as absolute scores.

(16) Ratings for the 1sg of the novel verb *anáxtape*

output	Actual rating	Predicted rating
wa'ánaxtape	9	0
anáwaxtape	9	8
awánaxtape	4	3
anáxtawape	6	0
ma'ánaxtape	0	0
anámaxtape	6	1
amánaxtape	5	1
anáxtamape	0	0

In general, we can isolate the following ways in which the actual ratings differed systematically from the predicted ratings: first, prefixed forms got globally higher ratings than predicted; this may be a result of having such a small database, and in particular of including so many verbs beginning with *o*-, *i*-, etc., in order to test those neighborhoods. It is possible that if we considered several thousand verbs, it would turn out that the vast majority of them actually prefix, and the predicted ratings for prefixes would be more similar to the actual ratings. (This would be parallel to the English or German case, where the regular, "default" pattern for past tenses is less common in the first few hundred most common verbs, but becomes the dominant pattern as we move down into the rarer words.)

A second difference between novel forms and the predictions based on existing forms was that certain phonological changes which happen regularly in existing words were not considered obligatory in novel words. For example, novel verbs beginning with *y*- sometimes sounded acceptable with the change *y*→*bl*, but they also tended to be acceptable with a prefixed *wa* (yielding *way*...). Similarly, the change of *k* to *c* after *i* is often "undone" when a *wa*- or *ma*- is inserted between the *i* and the *c* in existing words, but both *k* and *c* outputs were considered acceptable when I asked about these. In the case of *k* ~ *c* alternations, this may be due to a low level of productivity for the palatalization rule for this particular consultant. These intuitions probably also reflect an overall preference in language for rare or nonce forms to resist phonological alternations.

Another pattern was that the -*wa*- subject marker was preferred over the –*ma*– marker regardless of the phonological form of the root involved. One possible explanation of this would be that just like prefixation, the tendency to take *wa*- is in fact true in the overall lexicon, but it did not emerge in so small a file. However, in this case, I think that a more likely explanation is that in fact the choice of series I or series II subject marking is not a matter of arbitrary inflectional classes, but is rather due to semantic or syntactic factors. In particular, although there are some active verbs which take the Series II ("*ma*") markers, there are no transitive verbs which do. The novel verbs in this study were all presented in a transitive context in the frame sentence, so the Series I *wa-* was the most plausible choice.

Finally, there was no observed effect of more subtle phonological neighborhoods, such as the difference between *-pha* and *-kha*. For these, it is probable that either the differences are not as significant in the entire lexicon as it seemed to Boas and Deloria, or else the fact that I was only able to ask a single consultant about a rather small number of words meant that it was simply impossible to pick up on subtle differences like this.

4.2. *Correlation of ratings to predictions*

One way to test how well the actual ratings matched the predicted ratings is to perform a correlation between the two. I did this, and discovered that in fact there was a positive and significant correlation ($r = .333$, $p < .001$, where r=0 means no effect at all and r=1 means a perfect fit). However, this correlation is not especially strong, and is therefore not all that impressive.

One of the systematic differences between the predicted and actual ratings noted above was that *wa-* was generally preferred over *ma-* as the 1sg subject marker. One natural interpretation of this is that the difference between *wa-* and *ma-* is in fact a syntactic or semantic one. If this is true, then the choice of *wa-* or *ma-* would not be related at all to the phonological form of the novel verb, and the choice of which to use would be unrelated to the decision about where to infix. In order to test for this, I grouped the ratings for *wa-* and *ma-* forms, in order to see simply how good each word sounded with prefixed marking vs. infixed marking. For example, combining the *wa-* and *ma-* ratings for the novel verb *anáxtape* gave the following set of ratings:

(17) Ratings from (16), with *wa-* and *ma-* combined

output	Actual ratings	Predicted rating
*a*naxtape*	9	4.1
*ana*xtape*	15	9.25
*anaxta*pe*	6	0

I then recomputed the correlation between these combined predictions and the combined ratings, with a much better fit ($r = .538$, $p < .0001$). Figure 1 shows that there is indeed a trend for forms which are predicted to be better to in fact receive higher ratings. We can also notice that the computer has predicted a large number of 0's (the cluster of dots on the left side of the graph), so there may also be some problem with how the program is generalizing over infixes.

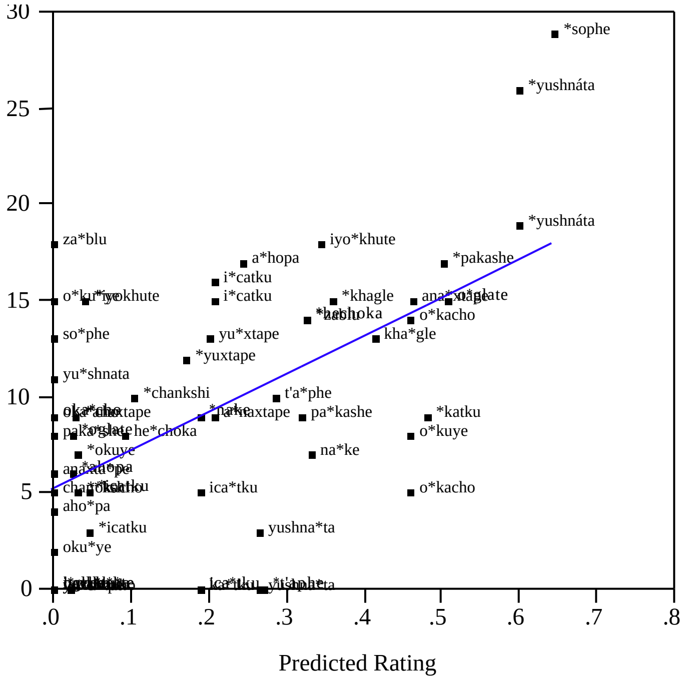

Figure 1: Correlation between actual ratings and predicted ratings
("*" = position of the subject marker)

5. CONCLUSION

Although this study did not confirm all of the subtle effects of phonology which Boas and Deloria suggest, it did find a definite effect of phonological form on the likelihood of novel verbs to infix. This demonstrates that in Lakhota, as in other languages, there are islands of reliablity for each morphological pattern. In particular, at least one Lakhota speaker feels that some novel words sound better with infixed subject marking than others, and this is a gradient effect. These intuitions could not possibly be based on any prior knowledge of the morphological decomposition of the novel words; rather, they are based on the behavior of similar words in the lexicon.

As I mentioned in section 2.3, part of the success of the computational model is due to the fact that it interprets morphemes as phonological neighborhoods, rather than having a notion of a "morpheme." At this point, it is natural to wonder how this approach could ever work — after all, don't we need to learn that the initial vowels are themselves locative prefixes, and that locative prefixes are placed before subject marking in the syntax? Here I would like to suggest that perhaps there are actually advantages to this morpheme-less approach in a system like Lakhota. The first relevant fact is that even in the existing lexicon, not all verbs which begin with *o*-, *i*-, or *a*- contain an obvious locative prefix — for example, the verb *ómna* 'smell' begins with an *o*- and takes an infixed subject marker (*ówamna* 'I smell'), but there is no unprefixed verb root *mna*, and the meaning 'smell' does not involve any obvious locative. So even in the existing lexicon, it is not clear that knowing morphemes will explain all cases — for some words we will still need to stipulate either the location of subject marking or else we will need to stipulate a morphemic analysis for the verb.

The second point to be made here is that for novel words, surely there is no evidence that there is actually a prefix on the word (since the consultant has not seen an unprefixed base for the word, and there is no meaning which could possibly hint that there is a locative meaning involved). So a strictly morpheme-based analysis actually misses the intuition that many words which participate in patterns do not strictly contain the "morpheme" semantically speaking, but they simply "look like they contain the morpheme." A similar phenomenon has been called "aggressive suffixation" by Hammond (1999), by which English words which contain the phonological material of a suffix also behave as if they are suffixed in the computation of stress. (For example, *honest* and *modest* have penultimate stress even though their final syllables are closed by two consonants, and this is claimed to be because the final syllable happens to look like the superlative suffix -*est*.) Therefore I believe that the brute force "phonological neighborhood detector" approach actually might model human behavior fairly well in cases where there is some ambiguity as to whether a word contains a prefix or not.

More generally, although the match to human judgments presented in section 4.2 is not a perfect fit, I believe that the results presented here do support the idea that the productivity of different morphological patterns is derived by looking at phonological properties of the roots involved. Experience with wug testing in other languages indicates that the results are much cleaner when there is more data (i.e., when you have collected judgments about many words containing each phonological neighborhood), and when you have data from a number of different consultants. Nevertheless, studies in other languages, like English and Italian, where it is easier to find dozens of consultants, have shown similar results, and this study therefore complements these other studies by demonstrating phonological effects on a typologically different pattern of morphological irregularity.

Appendix A. Actual and Predicted Ratings

(novel) 3sg. pres.	1sg. pres.	actual rating	predicted rating
ahópa	awáhopa	7	4.1
ahópa	wa’áhopa	6	0.4
ahópa	ahówapa	0	0.0
ahópa	amáhopa	10	0.7
ahópa	ma’áhopa	0	0.0
ahópa	ahómapa	4	0.0
anáxtape	wa’ánaxtape	9	0.4
anáxtape	anáwaxtape	9	7.9
anáxtape	awanaxtape	4	3.4
anáxtape	anaxtawape	6	0.0
anáxtape	ma’ánaxtape	0	0.1
anáxtape	anámaxtape	6	1.4
anáxtape	amanaxtape	5	0.7
anáxtape	anaxtamape	0	0.0
chankshí	machánkshi	10	2.0
chankshí	chanmákshi	5	0.0
héchokha	wahéchokha	9	4.1
héchokha	héwachokha	0	0.0
héchokha	héchowakha	0	0.0
héchokha	mahéchokha	5	2.4
héchokha	hémachokha	8	1.8
héchokha	héchomakha	0	0.0
icátku	waícatku	3	0.7
icátku	iwácatku	9	4.1
icátku	icáwatku	0	0.0
icátku	mícatku	0	0.1
icátku	imácatku	7	0.0
icátku	icámatku	5	3.7
iyókhute	wa’íyokhute	9	0.7
iyókhute	iwáyokhute	0	0.0
iyókhute	iyówakhute	9	4.1
iyókhute	iyókhuwate	0	0.0
iyókhute	maíyokhute	6	0.1
iyókhute	imáyokhute	0	0.0
iyókhute	iyómakhute	9	2.8
iyókhute	iyókhumate	0	0.0
katkú	wakátku	9	7.8
katkú	kawátku	0	0.0
katkú	makátku	0	1.8
katkú	kamátku	0	3.7
khaglé	wakhágle	5	5.4
khaglé	khawágle	7	8.3
khaglé	makhágle	10	1.8
khaglé	khamágle	6	0.0
náke	wanáke	9	2.2
náke	náwake	7	5.3
náke	manáke	0	1.5
náke	namáke	0	1.4
náke	manáwake	10	0.0
ogláte	waóglate	8	0.4
ogláte	owáglate	10	8.5
ogláte	ogláwate	0	0.0
ogláte	maóglate	0	0.0
ogláte	omáglate	5	1.6
okácho	waókacho	5	0.4
okácho	owákacho	9	7.6
okácho	okáwacho	0	0.0
okácho	maókacho	0	0.2
okácho	omákacho	4	1.6
okácho	okámacho	0	0.0
okácho	owákacho	6	7.6
okácho	okáwacho	6	0.0
okácho	waókacho	9	0.4
okácho	omákacho	0	1.6
okúye	waókuye	0	0.4
okúye	owákuye	3	7.6
okúye	okúwaye	2	0.0
okúye	owákuwaye	9	0.0
okúye	maókuye	7	0.2
okúye	omákuye	5	1.6
okúye	okumaye	0	0.0
okúye	omákuwaye	6	0.0
pakáshe	wapákashe	10	8.2
pakáshe	pawakashe	0	6.4
pakáshe	pakawashe	8	0.0
pakáshe	mapákashe	7	1.8
pakáshe	pamákashe	9	0.0
pakáshe	pakámashe	0	0.0
sophé	wawásophe	10	5.3
sophé	wasóphe	10	5.3
sophé	sowáphe	5	0.0
sophé	masophe	9	2.4
sophé	somaphe	8	0.0
t'aphé	wat’áwaphe	10	5.7
t'aphé	wat’áphe	0	5.4
t'aphé	mat’aphe	0	4.0
yushnáta	wayúshnata	4	2.0
yushnáta	blushnáta	5	8.6
yushnáta	yuwáshnata	9	0.0
yushnáta	mayúshnata	10	1.4
yushnáta	yumáshnata	2	0.0
yushnáta	yushnáwata	0	5.3
yushnáta (2)	wayúshnata	7	2.0
yushnáta (2)	blushnáta	10	8.6
yushnáta (2)	yuwáshnata	0	0.0
yushnáta (2)	mayúshnata	9	1.4
yushnáta (2)	yumáshnata	0	0.0
yushnáta (2)	yushnáwata	3	5.3
yuxtápe	wayúxtape	9	2.0
yuxtápe	yuwáxtape	6	4.0
yuxtápe	yuxtáwape	0	0.0
yuxtápe	mayúxtape	3	1.4
yuxtápe	yumáxtape	7	0.0
yuxtápe	yuxtámape	0	0.0
zablú	wazáblu	9	4.1
zablú	zawáblu	8	0.0
zablú	mazáblu	5	2.4
zablú	zamáblu	10	0.0

REFERENCES

ALBRIGHT, ADAM. 1998. Phonological subregularities in productive and non-productive verb classes: Evidence from Italian. MA Thesis, Los Angeles: UCLA.

ALBRIGHT, ADAM. and BRUCE HAYES. 1998. An Automated Learner for Phonology and Morphology. Ms. Los Angeles: UCLA.

BERKO, JEAN. 1958. The Child's Learning of English Morphology. *Word* 14(2-3), 150-177.

BOAS, FRANZ. and ELLA DELORIA. 1941. *Dakota Grammar*, Vol. 23 of *Memoirs of the National Academy of Sciences.* Washington: United States Government Printing Office.

BUECHEL, EUGENE. 1970. *A Dictionary of Teton Sioux*. Red Cloud Indian School, Inc.

BYBEE, JOAN, and DAN SLOBIN. 1982. Rules and Schemes in the Development and Use of the English Past Tense. *Language* 58(2), 265-289.

COWART, WAYNE. 1996. *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. SAGE Publications.

CROWHURST, MEGAN. 1998 *Um* infixation and prefixation in Toba Batak. *Language* 74(3), 590-604.

HAMMOND, MICHAEL. 1999 English stress & *cranberry* morphs. Paper presented at the 75th Annual Meeting of the Linguistic Society of America. Los Angeles.

MCCARTHY, JOHN and ALAN PRINCE. 1993. *Prosodic morphology 1*. RuCCS TR-3. New Brunswick, NJ: Rutgers Center for Cognitive Science.

MUNRO, PAMELA. 1989. Lakhota Verb List. Ms. Los Angeles: UCLA..

PRASADA, SANDEEP. and STEVEN PINKER. 1993. Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes* 8, 1-56.

ZUBIN, DAVID A., and KLAUS-MICHAEL KÖPCKE. 1981. Gender: A less than arbitrary category. Chicago Linguistic Society 17, 439-49.