Proceedings of the 39th Siouan and Caddoan Languages Conference

Held at Northeastern State University

Broken Arrow, Oklahoma

May 30-31, 2019

Editors

Ryan M. Kasak
Samiron Dutta
Justin T. McBride

Kasak, Ryan M., Samiron Dutta & Justin T. McBride (eds.). 2020. *Proceedings of the 39th Siouan and Caddoan Languages Conference*. Broken Arrow, OK: Northeastern State Univeristy Department of Languages and Literature.

This title can be downloaded at:
http://www.siouan.org/proceedings/sclc39/

ISSN: 2641-9904 (digital)

Typesetting: Ryan Kasak
Proofreading: Ryan Kasak, Justin T. McBride, & Samiron Dutta
Font: Times New Roman
Typesetting software: XeLaTeX & BibTeX

First printing, 7/31/2020

Contents

Acknowledgments

Many people helped to make SCLC 39 a success, and all of these individuals and organizations deserve thanks. This includes not only the presenters, but also the community members and scholars who attended their presentations and provided valuable feedback and encouragement. There were also many people working behind the scenes at Northeastern State University, on both the Tahlequah campus and the conference site at the Broken Arrow campus, and others among the university's various off-campus affiliates who lent their talents to the promotion, catering, lodging, and general logistics of the event for the benefit of all attendees. Finally, SCLC 39 would not have been possible without the very generous financial support of the Ho-Chunk Nation. *Ahó*!

Preface

SCLC 39 almost didn't happen. It was planned for a time when the Arkansas River watershed, particularly in Oklahoma, ended up experiencing some of the worst flooding to be witnessed in generations. In the week leading up to the event, conditions just miles from the conference site were bad enough to make national and even international headlines. Virtually all waterways in the region had spilled over their banks as a result of seemingly endless spring rain, overwhelming dams and locks, cutting off entire communities, displacing thousands of people from their homes, and leaving more than one hundred injured or dead. Many of the main highways throughout the area were closed to traffic—to say nothing of countless smaller roads that remained impassable for weeks after the others reopened—and flights in and out of Tulsa were delayed, rerouted, or simply canceled. After months of careful planning, it seemed that the conference was simply not meant to be. Suddenly, though, the rains diminished to light sprinkles and then to mere memories, and the sun, which had been entirely absent for days, emerged from parting clouds. Although a few of the attendees did still experience chaotic travel conditions to say the least, especially those who came by air, the conference once again became a possibility less than a day before the event was to begin. Things improved so dramatically in such a short time following the rain that I suspect some who drove in on the first day of SCLC 39 may even have wondered what all the fuss was about.

Despite the weather concerns, the conference itself was a success. There were unfortunately no Caddoan presentations in 2019, but attendees were treated to a wide variety of presentations on Catawba and various Missouri River, Mississippi Valley, and Ohio Valley Siouan languages. The Siouan and Caddoan Languages Conference has long been a means of bridging distances between, on one side, tribal community members, teachers, learners, and scholars and, on the other side, academics in the social sciences and humanities, many of whom are non-tribal; the 2019 event was no different. The 13 presentations included talks on formal grammatical and text analysis, comparative phonology, language attitudes and ideologies, language learning materials, and language program updates. In addition, there was a screening of the film *Umóⁿhoⁿ Íye: The Omaha Speaking*. By the end of the conference, we averaged 25 attendees a day. This figure included members from a half dozen or so Siouan and Caddoan tribes, including several Elders and one teenager whose first language is Ioway. Plus, we had one attendee who crossed an ocean to be present. As always, attendees also managed to catch up with one another and make new acquaintances. After all, SCLC has always been about community as much as anything else.

A word or two about this proceedings volume is in order here. While the Siouan and Caddoan Languages Conference will soon enter its fifth decade of existence, this volume is only the second one produced since the early 1990s. It follows the very successful proceedings volume that Ryan M. Kasak edited for SCLC 38, as well as his ultimate vision of resuming an annual record of quality SCLC scholarship. Indeed, the realization of this SCLC 39 volume is also thanks to Ryan's organizational and editorial efforts–not to mention his LaTeX expertise. As a member of the remaining editorial staff, including Samiron Dutta and myself, I would like to extend a hearty thanks to Ryan for all of his work on preserving and promoting the scholarship of the various contributors whose manuscripts are contained herein. *Wéwihnaⁿ eyáo, kóya*.

Finally, I would like to thank Edwin Ko for volunteering to organize SCLC 40 in 2020.

I wish you all the best of fortunes in that endeavor, and I hope for everyone's sake that you will not have to contend with any large-scale disasters in planning or hosting it. I'm only teasing, of course. I'm actually writing this foreword after having already attended the outstanding virtual conference made necessary as a result of the Covid-19 outbreak. Edwin did an absolutely amazing job on that one, despite perhaps some of the most unforeseen of all unforeseen circumstances. He has also volunteered to organize SCLC 41 in 2021, and he deserves all our thanks, support, and encouragement for that. Cheers to you, sir!

Justin T. McBride
Collinsville, Oklahoma

Another Catawba lexicon*

Samiron Dutta David Kaufman
Mt. Lebanon, PA *Santa Barbara, CA*

Abstract: Although Catawba has been studied over 100 years, from early 1800s to the mid-1900s, all serious investigations about Catawba started in the early 1900s. Currently, the vast majority of extant field notes can be traced to the last three speakers of Catawba. However, there are primary sources from the early 1800s that have not been analyzed. These primary sources are word lists, and they harken to a time when Catawba was more widely spoken and less impacted by the deleterious effects of language dormancy. This paper presents an additional lexicon of Catawba from earlier primary sources.

Keywords: Catawba, Barton, Mills, Gallatin, Proto-Siouan

1. Introduction

Catawba is a dormant language once spoken in southeastern United States, and is currently being revived Heinemann-Priest (2000). Catawba belongs to the Siouan-Catawban macro family. Of the Catawban languages, Catawba is the only one that has been extensively documented and studied.

Several scholars have compiled unpublished lexicons of the Catawba language. The table below summarizes the best-known lexicons.

Table 1: Summary of all known Catawba lexicons

Author	Year	Primary sources
Kathy Shea	1984	Siebert (1945a,b) Swadesh (1937) Susman (1935), Speck (1934) Michelson (1913) Gatschet (1900)
Paul Voorhis	1990s 2003	Speck (1934) Siebert (1945a,b)
Claudia Heinemann-Priest	1996	Lieber (1858) Dorsey (n.d.) Michelson (1913) Speck (1934)
Blair Rudes	2008	Siebert (1945a; 1945b, unpublished notes) McDavid (1941)

*The authors would like to thank the late Dr. Bob Rankin, the late Dr. Blair Rudes, and Kathy Shea for pioneering Catawban studies. Without their dedicated efforts, the authors of this paper could not study Catawba with the attention it deserves.

Other lesser known lexicons were compiled by Alexander Pickens, Raven McDavid, and Wes Taukchiray. Some investigators, like Shea and Priest, included a diverse scope of primary referenced materials. In contrast others, like Voorhis and Rudes, constructed lexicons from a much more limited scope.

Unfortunately, there is still a large volume of original field notes that have not yet been analyzed. Amongst them are older sources dating prior to the 1860s. These older references have historically been ignored, and to the authors' knowledge, these older references have never been analyzed. The focus of Catawba studies of the early 19th century was solely based on wordlist construction; thus, no grammatical material was ever captured until Oscar Lieber's publication in 1858. The objective of this paper is to construct a lexicon of older, less studied materials. Since these materials were not recorded by creditable linguists, they are necessarily of poor quality. Despite the challenges associated with analyzing older references, these sources may hold a lot of value in capturing archaic terms. As Siouan linguistic studies become more developed, it is hoped that these references will answer more questions on Catawba's ties to the Siouan language family for future scholars.

The relevant references used in this paper are:

- Benjamin Smith Barton (1798) published a set of 30 words in his publication *New Views of the Origin of the Tribes and Nations of America*. In his publication, Barton surveyed several Native American languages and attempted to construct family trees. His entire list, with analysis, appears in this paper. The first edition of his publication did not include any Catawba entries. Prior to his publication of the second edition, Barton noted, "[w]hile this edition was in the press, the arrival of two Katahba-Indians in Philadelphia afforded me an opportunity of collecting a small specimen of their language." (Barton 1798:22, Appendix). Therefore, only the second edition of Barton's book is referenced here.

- John L. Miller, President of the Ebenezer Academy, recorded at least 147 words. The original vocabulary by Miller has not been located. Gallatin published some or all of Miller's wordlist in 1836.

- Robert Mills (1826) published his *Statistics of South Carolina*. His publication provides an overview of the geography, history, and politics of South and North Carolina. It also provides Catawba words for the natural number set one through ten and eight words and phrases in Catawba. Mills's entire wordlist with analysis appears in this paper.

- Albert Gallatin (1836) published Catawba words compiled by John Miller and Benjamin Barton. Gallatin's entire wordlist minus number words, with analysis appears in this paper. Where items on Gallatin's wordlist clearly came from Barton's publication, we reference Barton's 1798 publication. Otherwise, items from Gallatin's wordlist is assumed to come from Miller's wordlist.

- James Owen Dorsey (n.d.) compared nearly 41 terms in Catawba against related Siouan languages. Dorsey did not perform any original field work in Catawba Country. However, looking at his list of Catawba terms, it is very clear that Dorsey had access to additional wordlists outside of Barton and Miller. Since this set of slip files are undated, it is not clear to the authors what this/these additional source/s are. Regardless, the authors make references

to these slip files for terms not previously attested in Shea (1984), Voorhis (n.d., 1992), or Rudes (2008a,b). It is likely that Dorsey had access to Gatschet's field notes. If so, there are vocabulary found in the slip files that were not published in Gatschet's (1900) grammar. Most of Gatschet's unpublished field notes are currently unanalyzed.

A casual observer may also find three additional references from the 1800s that are rarely referenced in more current research on Catawba. They are described below and largely omitted from this paper.

- Fredrick A. Porcher's papers (n.d., circa 1850s) contain portions of a handwritten draft of Oscar Lieber's vocabulary (published 1858). Since Lieber's vocabulary is among the best-known primary references in Catawba, the notes located in the Porcher papers is not analyzed in this paper. Claudia Heinemann (1996) and Blair Rudes (2008a,b) analyze Lieber's work in detail elsewhere. A future scholar may be interested in comparing the handwritten draft to Lieber's printed publication.

- Chamberlain (1888) was among the first to make the connection between Catawba and Woccon with other Siouan languages. He proposed a set of cognates based on wordlists available during his time. Because of recent scholarship (Rankin et al. 2015), future scholars can determine how accurate his observations were with Proto-Siouan. Chamberlain's pamphlet is now available in the Siouan Archives for a future scholar's reading pleasure. However, for the current study at hand, the authors only make brief mention of Chamberlain's work where appropriate. Chamberlain did not perform any original field studies and relied solely on words compiled by Miller and Barton, which are analyzed in this present paper.

- Albert Gallatin (1848) published 60 words in his *Hale's Indians of the North-West American, and Vocabularies of North America*. The sources for this list were the original Miller and Barton lists. However, there is one critical difference in this book, when compared to Barton's and what we believe to be Miller's lists. Gallatin attests that *gitlung* meant 'hair.' This is an error on Gallatin's notes. He meant to place this term in the Cherokee wordlist. Because this book is a repeat of information already provided by Miller and Barton, this book has not been used for any further analysis in this paper.

2. Methods

Wordlists from Barton (1798), Mills (1826), and Miller (as referenced in Gallatin 1836) were perused. Terms were analyzed by morphemes if possible. These terms were compared against lexicons previously compiled by Voorhis (n.d.), Shea (1984), and Rudes (2008a,b). When a morpheme was found to be previously undescribed, it was populated in Table 1 of this paper. All terms from Barton, Mills, and Miller's wordlists are available in Appendix A.

This paper defines "previously undescribed" as:

1. Morpheme or gloss attested by Barton, Mills, Miller does not appear on lexicons compiled by Shea (1984), Voorhis (n.d.), Rudes (2008a,b), or

2. If morpheme does appear on these aforementioned lexicons, Barton, Mills, and Miller are the sole sources for these morphemes.

A brief note of phonology and morphemic analysis is covered in Appendix B of this paper, largely based on previous scholarship.

3. Results and discussion

3.1. Limitations to this analysis

Analyzing these older references is compounded with several limitations endemic to Catawban language studies today. Mills, Miller, and Barton clearly had a difficult time hearing nasal vowels.

(1) *eesauh* (Gallatin 1836), *eswoa* (Barton 1798)
iswą (Shea 1984:187)
'river'

However, the nasal quality for this word was clearly attested by Oscar Lieber as "i-sw*ang*" (1858:7).

There are other cases where it is unclear if at one time a nasal vowel was optional or if the early investigators missed it. An example of this uncertainty is the morpheme for 'man'.

(2) *yabrecha* (Gallatin 1836)
yįį baritca (Shea 1984:177, referencing Speck 1934)
'man' (??)

When Miller attested this term, he did not give the first vowel (here, *a*) without any indication of a nasal quality. The Catawba term for man has been attested with a nasal vowel *yęę*, *yę*, *yįį* (see Shea 1984:177). However, before we dismiss Miller for making an error, observe what Gatschet attested in 1900:

(3) *ye' mi'-h-ra-re* (Shea 1984:177, referencing Gatschet 1900)
man superior-3S-PROX-IND
'chief, commander'

Gatschet, unlike Mills, Miller, and Barton, was trained in studying Native American languages. He did not record a nasal sound in the example above. In fact, Shea records several examples from Gatschet where the first vowel of this word was indeed oral, when later sources record an obligatory nasal vowel. By 1934, all attested words for 'man' in Frank Speck's notable publication *Catawba Texts* all record an obligatory nasal vowel (see Shea 1984:177 and Voorhis n.d.). It is, therefore, possible that the word for 'man' could have had an optional oral vowel sound in the early 1800s. Further evidence of this optional nasal sound is the fact that Miller had attested yakezuh, as opposed to *yaa kiča* for 'wife' and *yį kiča* for 'husband', for both husband and wife (see Appendix A, indices 47 and 48 of this paper)—thus, a difference between *yį* and *yaa* was not obvious to Miller.

The nature of vowels (short, long, or nasal) is conspicuously uncertain throughout the Catawba corpus. Gatschet, himself, indicated that the nasal was sometimes optional, but did not provide any further observation as to when the optional nasal sounds were allowed. Another example of this uncertainty is:

(4) *ɻti* (Shea 1984:188)
iti (Gatschet 1900:528)
'rock, stone'

The authors of this paper have an additional limitation. Most of the Catawba corpus has been unanalyzed and found in archives all over the United States. The authors here have made painstaking efforts to find and consolidate these unpublished references; but much of it is still untouched despite our efforts. It is quite likely that these items described in this paper as "archaic" may not actually be so. A great example of this is the entry for 'beaver'. Shea (1984) writes "[...] the vocabularies published by Gallatin are the only sources which cite a Catawba word for 'Beaver' *chaupee*, clearly a Siouan cognate." However, *čapi* was elicited by the late Chief Sam Blue *tonsi-tcop* or *chop* for 'beaver' (Pickens 1957). The authors of this paper are fully cognizant of the fact that they may be making the same error as well on other terms that have been attested on currently unanalyzed materials.

3.2. Catawba numbering system

The Catawba numerical system has been described elsewhere (Heinemann 1996). Thus, we will not spend much time on number related words here. Mills attested numbers from 1 through 10. Miller attested numbers 1-12, 20, 30, and 100. These entries are largely consistent with other scholarship.

However, there are three entries that are notable:

(5) *ne-purre* (Mills 1826)
nipare
'six' (error for 'two')

Mills clearly made an error here. The attested term is clearly for 'two' and not 'six'. The correct word for 'six' is *dípkuraa* (Rudes 2008b:22).

(6) *lubbosa* (Gallatin 1836)
nipisa (Mills 1826)
dawosa (Shea 1984:95)
'eight'

The number eight has been attested as *dawosa* (Shea 1984:95). As far as the authors are aware, *lubbosa* has not been attested elsewhere in the Catawba corpus and may indeed be an archaic term.

The third entry of note is as follows:

(7) *pat-chaw* (Mills 1826)
paca
wąča (Shea 1984:166)
'nine'

The attested term for 'nine' is *wąča*. Clearly either Mills did not hear the nasal quality of the verb or this term had an optional nasal vowel that later became obligatory.

3.3. Archaic phonemes

The wordlists compiled by Barton, Mills, and Miller have some typical phonemic alterations also observed in more recent Catawba corpus. Unexpected phonemes are discussed in this section.

The phoneme /l/, as attested in the entry *lubosa* 'eight' in example (6), is very elusive in Catawba. The authors are only aware of a few occurrences of this phoneme, including this one. /l/ was attested in *ílyii waasee*ˀ 'eel' (Michelson 1913:sheet 27). This word *ílyii* is likely a borrowing from English. The phoneme /l/ also appears in some Catawba proper names, Yanabe Yalengway and Willmannantanghkee, from the 1750s (Bauer 2016:50). This phoneme was also attested by Oscar Lieber in *ih-wohl-deh* 'young'. Otherwise, this phoneme is absent in Catawba. This phoneme was also attested in Woccon in the following words: *welka* for 'duck' and *palawa* for 'a turd'.

Another exotic phoneme of interest is the entry for 'white'.

(8) *saukchuh* (Gallatin 1836)
takči (Shea 1984:248)
'white'

The initial /s/ may be been an error by Miller, who recorded this term. As far as the authors are aware, 'white' has not been attested as *sakči* elsewhere. However, if a future scholar determines that Miller was not mistaken and this term was optionally *sakči* in addition to *takči*, he or she may investigate if this term is indeed related to the Proto-Siouan *asą́ą 'white'.

The opposite change is observed in the following example:

(9) *pot-tateerawah* (Gallatin 1836)
pase-taru-aa (Voorhis n.d.:77)
axe-large-NEG
'axe, hatchet'

In example (9), Miller heard a /t/ in the medial position of the first morpheme; however, in later sources, we see /s/.

The wordlists compiled by Barton, Mills, and Miller often do not have vowels phonemes that align with recent scholarship. However, there two observed noteworthy patterns that show some evidence of vowel shifts. There is a tendency when later attested Catawba requires /ę/, /e/, or /ee/, Mills/Miller/Barton's wordlist shows /a/ or /aa/. Examples of this are as follows (bold lettering by authors):

(10) *y**a**yeh* (Gallatin 1836)
*y**ę**-ye* (Voorhis n.d.:108)
man-GEN
'Indian'

(11) *w**a*** (Gallatin 1836)
*w**ee*** (Shea 1984:233)
'town'

(12) *y**ah**ho* (Gallatin 1836)
*y**e*** (Shea 1984:250)
'wind'

Alternatively, since Barton, Mills, and Miller were not trained linguists, it is possible that they heard phonemes /e/ or /ee/ and transcribed /a/ based on how an English speaker would say the name of the letter *a*. Similarly, they may have heard the phonemes /i/ or /ii/ and transcribed it as /e/ (Justin McBride, p.c., 2020). Before we can dismiss Miller from making an error, we should look to a related language, Woccon. Lawson (1709) attested *yauh-he* for 'Indian' and *yuh-hor* for 'wind'. Both words have a consistent vowel sound for the first vowel with Miller's wordlist. Thus, it may be possible that an archaic form of this term is *yaaye, with an oral, long /aa/ for 'Indian' and *yaa for 'wind'. We do not have a word in Woccon for town, city, or village.

Miller attested several color terms had once ended in a short /a/ or perhaps a /ə/ sound. Whereas, later attested Catawba requires a short /i/ sound.

(13) *houkch**uh*** (Gallatin 1836)
*hawukč**i*** (Shea 1984:56)
'black'

Another example of this possible vowel shift is in example (7), where Miller either records a short /a/ or /ə/ at the end of the utterance, and later attested Catawba uses /i/.

3.4. Previously undescribed morphemes

An analysis of wordlists constructed by Mills, Miller, and Barton did not reveal significant information about Catawba not previously described by Rudes (2008a,b), Voorhis (n.d.), or Shea (1984). However, the list below was constructed from Appendix A per methods described in Section 2 of this paper.

Table 2: List of glosses and morphemes previously unattested

Index	Morpheme or gloss	Comment
15	*nipisa*	Attested for the number eight.
26	*pičine hare i-ksa na-mu-aʔ*	Attested to mean 1000. Morphemes may mean 'I want 100 arms'.
27	*ni(n)yaa*	Attested for 'my daughter'. Mills accurately attested the pre-nasal sound before *yaa* 'woman'. Lieber also recorded this pre-nasal. However, the pre-nasal phoneme was no longer obligatory in later references (see Voorhis n.d.:13).
28, 31, 34, 44, 51, 52	*-da*	In older references, this morpheme was elicited for first person singular inalienable possession. More recent references elicit *-na*.
36	*ehopwe*	Archaic term for God. This word clearly has the potential modal suffix *-we*. Other morphemes for this entry are uncertain. It is likely that this word is related to the locative *haap* 'on.top'.

50, 54, 55A, 64, 75, 212	Indefinite possessor affix *i-*	The indefinite possessor is quite prolific in Speck's texts; however, it has not been described in published literature yet. This possessor is the inalienable possessor for nouns that do not identify who that noun belongs to. Also, this affix can be the subject to prefixing verbs when the subject is not identified.
52	*yaa-naa* woman-1s	Attested to mean 'sister'.
56	*himuu*	Attested to mean '[his] face'.
72, 78	*-piya*	This suffix is only used for hands and toes. Because this phoneme was never recorded by trained linguists, the exact form of this phoneme is unknown (ex, is it *-píyaa*, *-piiyá*, *-píiyaa*, etc.)
74	*i-ksa-pis* IDEF-arm-skin	Attested to mean 'fingernail'. Individual morphemes have all been attested before. However, this is the only time the authors are aware of this gloss being attested.
83	*haa* *di-haa* 1s-heart	Attested to mean 'heart'.
87	*wįsi wide'yu*	This entry was attested for 'warrior'. I have analyzed the morphemes here to be 'meat Shawnee'. However, morphemes are uncertain.
88	*ya-atehune-ee*	Attested to mean 'friend'.
90	*Kaupeyaweeracha* keepa yą wiirą-c-aʔ hole water to.burn-CAUS-SUB	Attested to mean 'kettle'. This entry's morphemes are uncertain. Best guess is provided.
107	*heakuh* *hiika*	Attested to mean 'light'.
113A	*yapha sohę* leaf fall-CONT 'leaf falling'	Attested to mean 'autumn'. Each morpheme has been previously attested. However, this gloss has not been previously attested. Interestingly, its form matches the Biloxi *snisníhi* which was attested to mean 'when the leaves begin to fall' (Dorsey n.d.).
113B	*woye* or *waye*	Attested to mean 'autumn'. Could be an error for 'winter'.
128	*háake*	Attested to mean 'lake'.
144	*yay*	Attested to mean 'oak'.
167	*siʔkąči*	Attested to mean 'red', attested previously as *siʔką* and *skę*. The second syllable *-či*, has not been attested with this color word. Further analysis is required.
169	*yą wi-h-aʔ* water resembles-STAT-SUB	Attested to mean 'blue'. Morphemes indicate 'water-like'. All of these morphemes have been previously attested. However, this gloss has not been attested.

180	*eʔ-h-ču wį-h-ree* to.dislike-3s-INTS to.resemble-3s-IND	Attested to mean ‘ugly. Morphemes are uncertain.
188	*wupchahaora*	Attested to mean ‘we’. Morphemes are uncertain.
190	*kera-arrera*	Attested to mean ‘they’. Morphemes are uncertain.
214	*nutátie*	Attested to mean ‘blanket’.

4. Conclusion

In a recent paper, de Reuse & Tefler undertook a similar survey to double check an old 1881 Nakoda wordlist, authored by Anna Barker, against current Nakoda speakers (2020). Their findings indicated current Nakoda speakers could re-elicit almost every word found in the wordlist by Anna Barker (Telfer 2020, p.c.). This paper’s findings show that much has changed in Catawba between surveys from pre-civil war era and the 1940s, when Frank Speck and Frank Siebert were actively studying this language. This rapid change over time may be because Catawba was already in the process of language death in the early 1800s.

Previously, Rudes (2003) identified several phonological changes that were occurring in Catawba between the mid 1800s to the mid-1900s. Rudes noted denasalization of phonemes /n/ and /m/ to /b/ and /d/, when they occurred before oral vowels. Examples of this phenomenon is *nurupi* (Gallatin 1836) to *duruubi* (Shea 1984:139) for ‘iron’; *nitemp* (Gallatin 1836) to *ditemp* (Shea 1984:45). Compare those words with *nų’ʔmąʔ* for ‘bead’, which did not denasalize (Rudes 2008b). Table 1 of this paper uncovers that certain terms like ehopwe for ‘God’, *híika* for ‘light’ fell into disuse. Morphemes such as *-piya* for ‘extremity’ and *-da* for first person alienable possession also fell into disuse over time.

Another curious observation is that the independent personal pronoun marker *-ta* or *-taʔ* was not elicited by Barton or Miller (see Indices 185 through 190). Lieber (1858) also did not elicit this pronoun marker. However, this marker is prolific in Siebert’s and Speck’s studies (Shea 1984:126).

The authors believe that examining other older, primary sources, other than the ones surveyed here, will uncover more archaic morphemes. An example of a previously documented morpheme is *aráta*, which is attested to mean ‘leader’ and a name or title of the famous Catawba, Chief Haigler (Rudes 2008a, referencing historical documents). This recorded term may be a cognate with proto-Siouan *(wa-)Róota which means ‘war’ (Rankin et al. 2015).

This paper identifies the following:

1. Almost two dozen morphemes that were not previously identified by Shea, Voorhis, and Rudes. They are described in Table 1.

2. Phoneme /l/ may have been used in older Catawba. Do note that /l/ was not a Proto-Siouan phoneme.

3. Possible sound shift from /*a/ or /*aa/ to /e/, /ee/, /ę/.

4. There is some evidence of English calques entering into Catawba. See Index 102 of Appendix A. The head-modifier configuration, which is typical in Catawba, would result in

nųti wíčaawa for 'moon'. This configuration was attested by Lieber (1858:7). However, Miller attested *wíčaawa nųti*, a configuration more similar to English. Miller either made an error or English was entering into Catawba by this time.

5. *Yaa*, or 'woman', may have been homophonous with *yaa-* for 'sister' (Appendix A, Index 52) and *-yaa* for 'daughter' (Appendix A, Index 27). Later attested Catawba shows a preference for *haču-* and *-inuwaa*, respectively and the former terms appear to have fallen into disuse. However, the data in Appendix A clearly shows that the term for 'daughter' has always been inalienably possessed and 'sister' has always been alienably possessed, regardless of the later attested terms or, what appears to be, archaic terms.

References

Barton, Benjamin Smith. 1798. *New views on the origin of the tribes and nations and North America*. Philadelphia, PA: John Bioren 2nd edn.

Bauer, Brook. 2016. *Being Catawba: The world of Sally New River, 1746-1840*. Chapell Hill, NC: University of North Carolina at Chapell Hill dissertation.

Chamberlain, Alexander F. 1888. *The Catawba language*. Toronto, ON: Imrie and Graham Printers.

de Reuse, Willem & Cory Tefler. 2020. Anna Barker's vocabulary of the Mountain Stoney Indian (1883-1886). Paper presented at the 40th Annual Siouan and Caddoan Langauges Conference. 23 May. Held remotely via Zoom.

Dorsey, James Owen. n.d. Comparative vocabular of Catawba, Tutelo, Biloxi, Hidatsa, etc. NAA MS 4800[82]. Washington, DC: National Anthropological Archives, Smithsonian Institution.

Gallatin, Albert. 1836. A synopsis of the Indian tribes within the United States east of the Rocky Mountains, and in the British and Russian possessions in North America. *American Antiquarian Society's Archaeologia Americana* 2. 1–422.

Gallatin, Albert. 1848. Hale's Indians of the north-west America, and the vocabularies of North America. *Transactions of the American Ethnographical Society* 2. xxiii–130.

Gatschet, Albert S. 1900. Grammatic sketch of the Catawba language. *American Anthropologist* 2(3). 527–549.

Heinemann, Claudia Y. 1996. *The Catawba language*. Rock Hill, SC: Winthrop University M.L.A. thesis.

Heinemann-Priest, Claudia Y. 2000. The applied linguist in Indian Country: Preservation and revival of the Catawba language. In Nicholas Ostler & Blair Rudes (eds.), *Endangered languages and literacy*, 47–52.

Lawson, John. 1709. *A new voyage to Carolina; containing the exact description and natural history of that country: Together with the present state thereof. And a journal of a thousand miles, travel'd thro' several nations of Indians. Giving a particular account of their customs, manners, &c*. London, UK: n.p.

Lieber, Oscar. 1858. *Vocabulary of the Catawba language*. Charleston, SC: James and Williams Printers.

McDavid, Raven I. 1941. Field notebooks. Ms. Philadelphia, PA: James M. Crawford Papers (Mss.Ms.Coll.66), American Philosophical Society.

Michelson, Truman. 1913. NAA MS2714 Catawba linguistic notes and texts collected by Truman Michelson. Washington, DC: National Anthropological Archives, Smithsonian Institution.

Mills, Robert. 1826. *Statistics of South Carolina including a view of its natural, civil, and military history, general and particular*. Charleston, SC: Hurlbut and Lloyd.

Pickens, Alexander. 1957. Contributions to Catawba ethnozoology. *Neighborhood Research* Spring. 1–5.

Porcher, Fredrick. n.d. Notes on the Catawba language.

Rankin, Robert L., Richard T. Carter, A. Wesley Jones, John E. Koontz, David S. Rood & Iren Hartmann (eds.). 2015. *Comparative Siouan dictionary*. Leipzig, Germany: Max Planck Institute for Evolutionary Anthropology. (Available online at http://csd.clld.org/).

Rudes, Blair A. 2003. Catawba phonemes. In Blair A. Rudes & David J. Costa (eds.), *Essays in Algonquian, Catawban and Iroquoian linguistics*, 217–254. Winnipeg, MB: Algonquian and Iroquoian Linguistics.

Rudes, Blair A. 2008a. Chapter 4: Morphology of verbs. Ms. Charlotte, NC: University of North Carolina at Charlotte.

Rudes, Blair A. 2008b. English-Catawba dictionary. Ms. Charlotte, NC: University of North Carolina at Charlotte.

Rudes, Blaire A. 2000. Resurrecting coastal Catawban: The reconstructed phonology and morphology of Woccon. *Journal of Southern Linguistics* 24(2). 228–244.

Shea, Kathleen. 1984. *A Catawba lexicon*. Lawrence, KS: University of Kansas M.A. thesis.

Siebert, Frank T., Jr. 1945a. Linguistic classification of Catawba: Part I. *International Journal of American Linguistics* 11(2). 100–104.

Siebert, Frank T., Jr. 1945b. Linguistic classification of Catawba: Part II. *International Journal of American Linguistics* 11(4). 211–218.

Speck, Frank. 1913. Some Catawba texts and folk-lore. *Journal of American Folklore* 26(102). 319–330.

Speck, Frank. 1934. *Catawba texts*. New York, NY: Columbia University Press.

Susman, Amelia. 1935. Catawba materials. Ms. Mss.497.3.B63c. Philadelphia, PA: American Council of Learned Societies Committee on Native American Languages, American Philosophical Society.

Swadesh, Morris. 1937. Catawba fieldnotes. Ms. Mss.497.3.B63c. Philadelphia, PA: American Council of Learned Societies Committee on Native American Languages, American Philosophical Society.

Voorhis, Paul. 1992. Catawba vocabulary from texts recorded by Frank Speck. Ms. Brandon, MB: Brandon University.

Voorhis, Paul. n.d. Catawba. Ms. Brandon, MB: Brandon University.

Young, Martha & Frank T. Siebert, Jr. 2003. Catawba onomastics. In Blair A. Rudes & David J. Costa (eds.), *Essays in Algonquian, Catawban and Iroquoian linguistics*, 271–274. Winnipeg, MB: Algonquian and Iroquoian Linguistics.

Appendix A

Table 3: Catawba lexicon from earlier sources

Index	Catawba-morpheme-English	Reference	Comments
1	*ne-po-ya* dapani 'one'	Mills (1826)	Entry is consistent with later attested Catawba. See Shea (1984:71).
2	*dupunna* dapani 'one'	Gallatin (1836:358)	Entry is consistent with later attested Catawba. See Shea (1984:71).
3	*nau-pa-re* nąpre 'two'	Mills (1826)	Entry is consistent with later attested Catawba. See Shea (1984:236).
4	*naperra* nąpre 'two'	Gallatin (1836:358)	Entry is consistent with later attested Catawba. See Shea (1984:236).
5	*no-mon-da* naamnaʔ 'three'	Mills (1826)	Entry is consistent with later attested Catawba. See Shea (1984:229).
6	*namunda* naamnaʔ 'three'	Gallatin (1836:358)	Entry is consistent with later attested Catawba. See Shea (1984:229).
7	*purree-purree* parpare 'four'	Mills (1826)	Entry is consistent with later attested Catawba. See Shea (1984:107).
8	*purrepurra* parpare 'four'	Gallatin (1836:359)	Entry is consistent with later attested Catawba. See Shea (1984:107).
9	*puc-tree* paktire 'five'	Mills (1826)	Entry is consistent with later attested Catawba. See Shea (1984:103).
10	*pukte-arra* paktire 'five'	Gallatin (1836:359)	Entry is consistent with later attested Catawba. See Shea (1984:103).
11	*ne-purree* depkara 'six'	Mills (1826)	Mills may have made an error here for 'two'. See Indices 3 and 4.
12	*dip-kurra* depkara 'six'	Gallatin (1836:359)	Entry is consistent with later attested Catawba. See Shea (1984:201).

13	*was-sin-nee* wasignure 'seven'	Mills (1826)	Entry is consistent with later attested Catawba. See Shea (1984:195).
14	*wassin-e-u* wasignure 'seven'	Gallatin (1836:360)	Entry is consistent with later attested Catawba. See Shea (1984:195).
15	*ne-pis-saw* dawosa 'eight'	Mills (1826)	Mills clearly attested an ancient form of this number word. The Woccon is *nupsau* (Lawson 1709). This number word has been previously recorded as dawosa (Shea 1984:95). However, Oscar Lieber too recorded *na-po-sah* for 'eight' (Lieber 1858:9).
16	*lubbosa* dawosa 'eight'	Gallatin (1836:360)	Miller records a possible older form of this word, with the elusive phoneme /l/.
17	*pat-chaw* wąča 'nine	Mills (1826)	Mills may have recorded an archaic form of this word.
18	*wunchah* wąča 'nine'	Gallatin (1836:360)	Entry is consistent with later attested Catawba. See Shea (1984:166).
19	*pitch-in-nee* pičine 'ten'	Mills (1826)	Entry is consistent with later attested Catawba. See Shea (1984:223).
20	*pechuna* pičine 'ten'	Gallatin (1836:361)	Entry is consistent with later attested Catawba. See Shea (1984:223).
21	*pechuna dupa hoksu* pičine dapę haksa 'eleven'	Gallatin (1836:361)	Entry is consistent with later attested Catawba. See Shea (1984:168).
22	*pechuna napurra hoksu* pičine nąpre haksa 'twelve'	Gallatin (1836:361)	Entry is consistent with later attested Catawba. See Gatschet (1900:535).
23	*pechuna napurra* pičine nąpre 'twenty'	Gallatin (1836:362)	Entry is consistent with later attested Catawba. See Gatschet (1900:535).
24	*pechuna namunda* pičine naamnaʔ 'thirty'	Gallatin (1836:362)	Entry is consistent with later attested Catawba. See Gatschet (1900).
25	*pechu-hahruh* pičine hare 'one hundred'	Gallatin (1836:362)	Entry is consistent with later attested Catawba. See Shea (1984:169).

26	*pechuhruh eeksuh namohah* pičine ha-re i-ksa na-mu-aʔ ten STAT-IND IDEF-arm 1s-want-SUB 'one thousand'	Gallatin (1836:363)	This form for 'one thousand' has not been attested. Gatschet attested *pičine hare pičine* Gatschet (1900:535). Although morphemes are uncertain here, this utterance may mean 'I want 100 arms'.
27	*non-yaw* *ni(n)yaa* ni-(n)yaa my-woman 'daughter'	Mills (1826)	Later attested Catawba for 'my daughter' is: *ni'nuwaa* my.daughter 'my daughter' Shea (1984:84) Mills's entry is interesting for the following reasons: 1. Mills (1826) records a nasal quality to the word 'woman'. Lieber also recorded this nasality as *n'-yah* (1858:7). However, this nasal sound preceding the word for 'woman' was not elicited in more recent references, it was elicited, instead, as *yaa* or *iyaa* (Voorhis n.d.:13, 104). 2. The fact that *-nuwaa* was not elicited for 'daughter' indicates that 'woman' was used for this kinship term or Mills was mistaken.
28	*cow-re-dha-har-ree* kuri-na ha-ree son-my STAT-IND 'my son'	Mills (1826)	This entry shows that the first-person alienable possessive marker could have been *-dha* and later attested Catawba may have nasalized it to *-na*.
29	*cow-sin-nee-wi-ra* kurį-na wi-ree son-1s to.resemble-IND 'resembles my son' a fine boy	Mills (1826)	Entry is consistent with later Catawba. See Shea (1984:209).

30	*ya-wee-can-nee* yaa wikąnee woman warm-INT 'is the woman warm?' a beautiful girl	Mills (1826)	*Wiką* can either be an archaic root for 'be.beautiful' or Mills made an error. It appears that his informant was asking 'is the woman warm?'
31	*cow-re-dhagh* kuri-na son-my 'a baby'	Mills (1826)	See index 28. Mills's informant was most likely saying 'my son'.
32	*ne-mough-sa-ragh* ne-mosa-ree I-love-IND 'I love you'	Mills (1826)	Entry is consistent with later attested Catawba. See Shea (1984:151).
33	*yon-e-go-jau* yaa ni-kičaa woman 1s.own 'my wife'	Mills (1826)	Interesting use of 1s. In later attested Catawba, 'I own' is *dikįčaa*, and Lieber reports that 'my wife' is *yaa-de-ke-tcheh*, which is consistent with Rudes's lexicon. However, Mills records that the first person singular may once have been *ni-*. Sally Brown, much later attested that 'my wife' was *yaa kįcaanaa* (Rudes 2008a). Truman Michelson also attested that /č/ is allophonic with /j/ (Michelson 1913).
34	*borough-hend-ha* bara-hę-daa brother-??-my 'my brother'	Mills (1826)	Mills records the archaic form of the first-person singular alienable possession marker, *-da*. It is not clear to the authors what the *hę* means.
35	*cunree-har-ree-yaw-ee* kurii-h-ree yaa-ee good-3s-IND woman-small 'the girl that I love'	Mills (1826)	See Shea (1984:85). *widee e*=deer calf.
36	*ehopweh* 'God'	Gallatin (1836:307)	This is an archaic word for god. God is *wariwe* in later attested Catawba, or 'man not dying'. Compare with Shea (1984:115).

37	*yahwerejeh* yį wirą-c-aʔ man to.die-CAUS-SUB 'wicked spirit'	Gallatin (1836:307)	Miller did not hear the nasal quality. However, the nasalization does appear in Lieber's (1858) publication. Michelson (1913:sheets 76/151) attests that the verb phrase's initial syllable was omitted, recorded as *yę rąčé*.
38	*yabrecha* ya breča man ?? 'man'	Gallatin (1836:308)	In more recent field notes, the word for man has been elicited as *yį* or *yę*. Either the recorder, when recording this entry, did not hear the nasal quality to man or the nasal quality was recently innovated in Catawba. The elicited phrase for man here is consistent with Lawson's elicited phrase in Woccon *yauh-* Rudes (2000). The morpheme *brecha* is unclear in meaning. It was later attested in Speck's notes (Voorhis n.d.:109).
39	*eeyauh* yaa 'woman'	Barton (1798:104)	Entry is consistent with later attested Catawba. See Shea (1984:252).
40	*eechahuh* yį ča h-a? man small 'boy'	Gallatin (1836:308)	Entry is consistent with later attested Catawba. Note Michelson attests *hįca* for 'boy' (1913:sheet 64/151).
41	*yahwachahuh* yaa wa ča he woman-??-small-?? 'girl'	Gallatin (1836:309)	Entry is consistent with later attested Catawba. See Shea (1984:204) for 'small'.
42	*yeenturawa* yį taro-aa man large-NEG 'child, infant'	Gallatin (1836:309)	Entry is consistent with later attested Catawba. See Shea (1984:204) for 'small'.
43	*yahmosa* yį musa man to.love 'father'	Gallatin (1836:309)	Here, *musa* 'to love' is used to mean 'beloved man'. Note that Miller did not record the nasal quality. See Shea (1984:151).

44	*nenedau* nane-na father-1s 'my father' father	Barton (1798:102)	Entry is consistent with later attested Catawba. See Shea (1984:98). Note archaic possessive marker.
45	*yaxu* yaksu 'mother'	Gallatin (1836:310)	Entry is consistent with later attested Catawba. See Shea (1984:161).
46	*checheendau* činčin-na mother-1s 'my mother' mother	Barton (1798:102)	Entry is consistent with later attested Catawba. See Shea (1984:168). Note archaic possessive marker.
47	*yakezuh* yį kiča 'husband'	Gallatin (1836:310)	Note the replacement of *č* with *z*. Also note Miller could not differential between *yį* (for 'husband') and *yaa* (for 'wife'), indicating that *ya may have been the same gloss for 'husband' and 'wife'. Otherwise, entry is consistent with later attested Catawba. See Shea (1984:211). Compare with Index 48.
48	*yakezuh* yaa kiča 'wife'	Gallatin (1836:310)	Note the replacement of *č* with *z*. Also note Miller could not differential between *yį* (for 'husband') and *yaa* (for 'wife'), indicating that *ya may have been the same gloss for 'husband' and 'wife'. Otherwise, entry is consistent with later attested Catawba. See Shea (1984:211). Compare with Index 47.
49	*koorewa* kuriiwaa son+our 'son'	Gallatin (1836:311)	Entry is consistent with later attested Catawba. See Shea (1984:209).
50	*enewah* i-inuwaa IDEF.daughter 'daughter'	Gallatin (1836:311)	Entry is consistent with later attested Catawba. See Shea (1984:84).

51	*murraundau* mbara-na brother-my 'brother'	Barton (1798:103)	Note that Miller too records the archaic first person alienable possessive marker. See Shea (1984:62).
52	*yadah* yaa-naa woman-my 'sister'	Gallatin (1836:312)	'Sister', in later attested Catawba is *hacu* and is alienably possessed (see Shea 1984:200). However, Miller's informant may have said 'my woman' (see Shea 1984:252).
53	*yayeh* yįye man+GEN 'Indian'	Gallatin (1836:312)	Woccon: *yauh-he* (Lawson 1709). Note the absence of the nasal quality in 'man' (see Shea 1984:137). Lieber also did not attest a nasal quality in his publication. His finding is consistent with Gallatin (see entry for 'Indian' (Lieber 1858:7)).
54	*iska* *-iskąʔ* 'head'	Gallatin (1836:312)	Entry is consistent with later attested Catawba. See Shea (1984:126).
55A	*iskonsa* i-ską-si IDEF-head-hair 'hair'	Gallatin (1836:313)	Compare this entry with index 54. Miller was able to hear and record the nasal quality in this entry, however, he missed it in the previous entry.
55B	*gitlung* 'hair'	Gallatin (1848:94)	This entry was an error in Gallatin's book. He meant to include this term in the Cherokee wordlist.
56	*heemoh* hi-nen 3s-face 'face'	Gallatin (1836:313)	This term is either an error on Miller's part or indeed an archaic term for 'face'. Compare with Shea (1984:96).
57	*eetaup* i-taap IDEF-forehead 'forehead'	Gallatin (1836:313)	Entry is consistent with later attested Catawba. See Shea (1984:106).
58	*netaup* ni-taap 1s-forehead 'my forehead' forehead	Barton (1798:109)	Entry is consistent with later attested Catawba. See Shea (1984:106).

59	*doxu* *duksuuʔ* 'ear'	Gallatin (1836:314)	Entry is consistent with later attested Catawba. See Shea (1984:93).
60	*nocksoo* nu-ksu 1s-ear 'my ear' ear	Barton (1798:109)	Entry is consistent with later attested Catawba. See Shea (1984:93).
61	*heetooh* hi̱tuuʔ 'his-eye' eye	Gallatin (1836:314)	Entry is consistent with later attested Catawba. See Shea (1984:96).
62	*neetooh* n-i̱tuuʔ 1s-eye 'my eye' eye	Barton (1798:108)	Entry is consistent with later attested Catawba. See Shea (1984:96).
63	*eepeesooh* i-pi̱suuʔ 'nose'	Barton (1798:107)	Note that Barton did not record the nasal quality. Otherwise, entry is consistent with later attested Catawba. See Shea (1984:166). The Proto-Siouan-Catawban cognate is *pa. The proposed Woccon cognate is *poppe* which also lacks the nasal quality seen in Catawba (Rudes 2000).
64	*esomo* i-sumuuʔ IDEF-mouth	Gallatin (1836:315)	Entry is consistent with later attested Catawba. See Shea (1984:162).
65	*heesoomoseh* hi-sumuuʔ-se 3s-mouth-point 'tongue'	Gallatin (1836:315)	Entry is consistent with later attested Catawba. See Shea (1984:162).
66	*neesoomoseh* ni-sumuuʔ-se 1s-mouth-point 'tongue' my tongue	Barton (1798:111)	Entry is consistent with later attested Catawba. See Shea (1984:162).
67	*heeaup* hi-yaap 'his tooth' tooth	Gallatin (1836:315)	Entry is consistent with later attested Catawba. See Shea (1984:232).

68	*neeaup* ni-yaap 1s-tooth 'my tooth' tooth	Barton (1798:110)	Entry is consistent with later attested Catawba. See Shea (1984:232).
69	*esomoesa* yi-sumuuʔ-see your-mouth-hair 'beard'	Gallatin (1836:316)	Entry is consistent with later attested Catawba. See Shea (1984).
70	*edut* *-idat* 'neck'	Gallatin (1836:316)	Entry is consistent with later attested Catawba. See Shea (1984:164).
71	*eeksuh* *-iksa* 'arm'	Gallatin (1836:316)	Entry is consistent with later attested Catawba. See Shea (1984:122).
72	*ecksapeeah* i-ksa-piya 'hand'	Gallatin (1836:317)	Miller reveals an archaic form for 'extremity'. See index 78. The same suffix *peeah* is used there too. This *peeah* suffix appears in Lieber's publication (see 'hand' in Lieber 1858:7).
73	*eekseaah* i-ksa icaa? 'fingers'	Gallatin (1836:317)	In this entry Miller reveals that the informant did not use the later attested Catawba suffix for 'extremity', *icaa?* for this word. However, the *icaa?* suffix appears in Index 80, 'toes'. Inference is *-iksa* can mean 'hand', 'arm', and 'fingers'.
74	*eeksapis* i-ksa pis hand-skin 'nails'	Gallatin (1836:317)	This entry is not found in the more well-known lexicons of Catawba. See Shea (1984:122).
75	*eehageo* i-haak-yuu IDEF-body-flesh 'body'	Gallatin (1836:318)	Entry is consistent with later attested Catawba. See Shea (1984:59).
76	*eepah* i-pą 'belly'	Gallatin (1836:318)	Entry is consistent with later attested Catawba. See Shea (1984:55).
77	No-Entry 'leg'	Gallatin (1836:318)	Catawba does not differentiate between 'leg' and 'foot'. Miller did not record an entry.

78	*hepapeeah* h-ịpa-piya 3s-leg-extremity 'feet'	Gallatin (1836:319)	Entry is consistent with later attested Catawba. See Shea (1984:105).
79	*nepapee'ah* n-ịpa-piya 1s-leg-extremity 'foot' my foot	Barton (1798:114)	Entry is consistent with later attested Catawba. See Shea (1984:105).
80	*epuhyetah* i-ịpa-ičaaʔ IDEF-foot-extremity 'toes'	Gallatin (1836:319)	Entry is consistent with later attested Catawba. See Shea (1984:105).
81	*heposaup* hi-pasaap 3s-shin 'his shin' bone	Gallatin (1836:319)	Miller's informant likely said *hị'pasaap*, or '(his) shin' (Rudes 2008b:98).
82	*nosaup* *nu-saap* 1s-bone 'my bone' bone	Barton (1798:89)	Entry is consistent with later attested Catawba. See Shea (1984:59).
83	*deehauh* di-haa 'heart'	Barton (1798:117)	Later attested Catawba has 'heart' as *-imę*. See Shea (1984:127). This entry may be an archaic word.
84	*eet* iit 'blood'	Barton (1798:116)	Entry is consistent with later attested Catawba. See Shea (1984:58).
85	*wa* wee 'town, village'	Gallatin (1836:320)	Entry is consistent with later attested Catawba. Note vowel shift from *wa* to *wee*. See Shea (1984:233).
86	*yahmeraee* yę miirá 'chief'	Gallatin (1836:321)	Entry is consistent with later attested Catawba. Note the absence of nasal quality. See Shea (1984:69).
87	*weeseeweedheu* wịsi widé-yu Shawnee cow-flesh meat Shawnee 'warrior'	Gallatin (1836:321)	Morphemes are uncertain. The authors provided their best guess. See Young & Siebert (2003:272) for possible context with Shawnee.

			According to Speck, *wịsi* is a "fierce monster in Catawba mythology" (Speck 1913:326). Although, Speck had later reanalyzed as 'many people' in his *Catawba Texts* (Speck 1934:91). It is possible that this term is a reference to this mythical monster.
88	*ya-atehune-ee* yę atehuneʔ-aʔ man friend-SUB 'friend'	Gallatin (1836:321)	Lieber attested the following for 'enemy': *ate-koo-ni-wah* (1858:8) atekuni-w-aa-∅ friend-??-NEG-SUB It is likely that the morpheme *kuni* 'good' is used in this entry's construction.
89	*sook* suk 'house, hut'	Barton (1798:91)	This entry is consistent with later attested Catawba. See Shea (1984:131).
90	*kaupeyaweeracha* keepa yą wiirą-c-aʔ hole water to.burn-CAUS-SUB 'kettle'	Gallatin (1836:322)	This term was not attested in latter references. Morphemes are uncertain.
91	*wah* wą 'arrow'	Gallatin (1836:322)	Entry is consistent with later attested Catawba. Miller missed the nasal quality. See Shea (1984:47).
92	*eecheka* *ičika* 'bow'	Gallatin (1836:323)	Entry is consistent with later attested Catawba. See Shea (1984:60).
93	*pot-tateerawah* pase-taru-aa axe-large-NEG small axe 'axe, hatchet'	Gallatin (1836:323)	Latter references attested *pvseʔ* for 'axe' (Voorhis n.d.:77). It is not clear if Miller misheard *pot-ta* for *pase* or if the former is an archaic expression for 'axe'.
94	*seepah* sępa 'knife'	Gallatin (1836:323)	Miller missed the nasal quality of the first vowel. Otherwise entry is consistent with later attested Catawba. See Shea (1984:142).

95	*dupomorya* dapa amáa-yaʔ something watercraft 'canoe, boat'	Gallatin (1836:324)	Entry is consistent with later attested Catawba. See Rudes (2008b:18). Recorder misheard an *r*.
96	*weeda* witée 'Indian shoes'	Gallatin (1836:324)	Entry is consistent with later attested Catawba. See Shea (1984:197).
97	*koostauh* kustą 'bread'	Barton (1798:88)	Entry is consistent with later attested Catawba. Recorder missed the nasal quality. See Shea (1984:51).
98	*wahmezu* wamsu 'pipe, calumet'	Gallatin (1836:325)	Entry is consistent with later attested Catawba. Note that sometimes *s* can have an allophone *z*. See Shea (1984:180).
99	*opah* įpa 'tobacco'	Gallatin (1836:325)	Entry is consistent with later attested Catawba. See Shea (1984:231).
100	*wahpeeh* wáapit 'sky, heaven'	Gallatin (1836:325)	Entry is consistent with later attested Catawba. See Lieber (1858:7). Note recorder did not hear the final stop consonant. Also, Rudes (2000:240) postulated that this word either Woccon or Catawba metathesized the middle two syllables. Observe the following: PS: *wiirą 'moon, sun, orb' (Rankin et al. 2015) W: *witáapiree (Rudes 2000:234) C: *wáapit* (Rudes 2000:234) If these words are indeed cognates, the proto-Siouan long /ii/ was reduced to a short /i/ and Catawba might have metathesized the middle two syllables. This, the Comparative Siouan Dictionary, offers insight to Rudes's question as to which of the two Catawban languages metathesized these syllables.

101	*nootéeh* nụti 'sun'	Barton (1798:119)	Entry is consistent with later attested Catawba. See Shea (1984:217).
102	*weechawa nooteeh* wícaawa nụti night sun 'moon'	Gallatin (1836:326)	Entry is consistent with later attested Catawba. See Shea (1984:160). In Catawba, 'moon' and 'sun' have the same gloss. Please note that Catawba noun phrases are head-modifier configuration. This gloss appears to be an English calque for 'night sun'. Lieber (1858:7) attests the head-modifier configuration for this gloss.
103	*nootéeh* nụti 'moon'	Barton (1798:121)	Entry is consistent with later attested Catawba. See Shea (1984:160).
104	*wahpeeknu* wáápitnu 'star'	Gallatin (1836:326)	Entry is consistent with later attested Catawba. Note that recorder did not hear the final stop, compare with 'sky, heaven' in index 101. Also see Shea (1984:214).
105	*yahbra* 'day'	Gallatin (1836:327)	Entry is consistent with later attested Catawba. Neither Rudes (2008a,b) nor the authors know what the morpheme *-bra* is in this context. Swadesh attested that this word means 'daylight' (Shea 1984:84).
106	*weechawa* wíčaawa 'night'	Gallatin (1836:327)	Entry is consistent with later attested Catawba. Frank Speck had this word varying with *hiičááwa*. See Shea (1984:165).
107	*heakuh* 'light'	Gallatin (1836:327)	Later attested Catawba uses *hayak*. Either this entry is an error or this is indeed an archaic term. *Hayak* was also used for 'lightning' in this paper in Index 116.

108	*weechaupku* wičɔ-k-aʔ night-HAB-SUB 'darkness'	Gallatin (1836:328)	Entry is consistent with later attested Catawba. Note the coalescence from *wičawa* to *wičɔ*. This entry uses the habitative suffix and the particle suffix. See Shea (1984:165).
109	*yahwup* yawap 'morning'	Gallatin (1836:328)	Entry is consistent with later attested Catawba. Note that Shea (1984:161) attests to a coalesced word *yɔɔb*.
110	*weechawa rare* night-PROX-IND 'evening'	Gallatin (1836:328)	Entry is consistent with later attested Catawba. See Shea (1984:165).
111	*yahrunkquechuh* yaarą kica summer own 'spring'	Gallatin (1836:329)	Rudes (2008b) documents 'spring' as *yáarirąʔ*. Alternatively, roots to this gloss might be: *yaarą ki-ca* summer this-DIM
112	*yahruh* yaarą 'summer'	Gallatin (1836:329)	Entry is consistent with later attested Catawba. See Shea (1984:217).
113A	*yup-hasohuh* yapha sohę leaf fall+CONT 'autumn'	Gallatin (1836:329)	*Wasiiʔ* was documented by Shea (1984:97). This gloss may be previously unattested.
113B	*woye*, *waye* 'autumn'	Dorsey (n.d.)	This gloss may be previously unattested. However, this gloss matches 'winter'. See Index 114.
114	*weeyah* wayą 'winter'	Gallatin (1836:330)	Archaic term for winter (Rudes 2008b). Proto-Siouan form is *wáarą 'winter'. Later attested Catawba term is *wayą*, as recorded by Shea (1984:251).
115	*yahho* ye 'wind'	Gallatin (1836:330)	Later attested Catawba is *yéhuu* 'wind is blowing' (Rudes 2008b). The archaic root of the word is closer to Woccon, *yuh-hor*. Note that archaic Catawba and Woccon attest to *yaa*/*yuu*.

116	h*iunk-hiunk-huh* hayak∼hayak-h-aʔ light-REDUP-STAT-SUB ‘lightning’	Gallatin (1836:330)Entry is consistent with later attested Catawba. See Shea (1984:147)	*Hayak* by itself means ‘to lighten’. However, when reduplicated, it means ‘lightning’. It is interesting that Miller heard a nasal quality.
117	*tere-re-hera* tiririhiree ‘thunder’	Gallatin (1836:331)	Entry is consistent with later attested Catawba. See Shea (1984:230).
118	*ookso’reh’* uksuʔ ‘rain’	Barton (1798:123)	Entry is consistent with later attested Catawba. See Shea (1984:184).
119	*wauh* waa ‘snow’	Barton (1798:124)	Entry is consistent with later attested Catawba. See Shea (1984:206).
120	*wauh sah* wasa ‘hail’	Gallatin (1836:332)	Entry is consistent with later attested Catawba. See Shea (1984:122).
121	*epee* ịịpi ‘fire’	Barton (1798:132)	Recorder did not hear nasal quality. Otherwise entry is consistent with later attested Catawba. See Shea (1984:102).
122	*eyau* yą ‘water	Barton (1798:131)	Entry is consistent with later attested Catawba. See Shea (1984:244).
123	*moha* mọha ‘ice’	Gallatin (1836:333)	Recorder did not hear nasal quality. Otherwise entry is consistent with later attested Catawba. See Shea (1984:136).
124	*munno* manuu ‘earth, land’	Gallatin (1836:333)	Entry is consistent with later attested Catawba. See Shea (1984:120).
125	*iswasekera* iswą sikirii river nasty ‘sea’	Gallatin (1836:333)	Entry is consistent with later attested Catawba. See Rudes (2008b).
126	*eesauh* iswą ‘river’	Gallatin (1836:334)	Entry is consistent with later attested Catawba. Notice lack of nasalization. See Shea (1984:187).
127	*eswoa, e-swo-a* iswą ‘river’	Barton (1798:81)	Entry is consistent with later attested Catawba. See Shea (1984:187).
128	*haukhe* hááke ‘lake’	Gallatin (1836:334)	Not previously attested.

129	*uhwah* 'valley'	Gallatin (1836:334)	Chamberlain (1888) and Heinemann (1996) offer plausible origins of this word. See Shea (1984:239).
130	*sookterrowa* suk-taro-aa mountain-large-NEG 'hill'	Gallatin (1836:335)	Entry is consistent with later attested Catawba. See Shea (1984:161).
131	*sook-taro* suk-taro mountain-large 'hill'	Barton (1798:79)	Entry is consistent with later attested Catawba. The negative suffix is apparently not required to differentiate between 'mountain' and 'hill]. See Shea (1984:161).
132	*sookterro* suk-taro mountain-large 'mountain'	Gallatin (1836:335)	Entry is consistent with later attested Catawba. See Shea (1984:161).
133	*suck, táro* suk-taro mountain-large 'mountain'	Barton (1798:77)	Entry is consistent with later attested Catawba. See Shea (1984:161). What Barton might be suggesting is that word accents are morphemic.
134	*sauwa* saawę 'island'	Gallatin (1836:335)	Entry is consistent with later attested Catawba. See Shea (1984:140). Miller missed the nasal quality of the last vowel.
135	*eedee* ịti 'stone, rock'	Gallatin (1836:336)	Entry is consistent with later attested Catawba. See Shea (1984:188).
136	*noropeweyeh* nuruubi wiya iron wire 'copper'	Gallatin (1836:336)	Word is not attested in Voorhis or Shea. See Rudes (2008b:83).
137	*norope-ee* duruubi-DIM 'iron'	Gallatin (1836:336)	Entry is consistent with later attested Catawba. Note that archaic Catawba once nasalized the initial /d/. See Shea (1984:139).
138	*koos* kus 'maize'	Gallatin (1836:337)	Entry is consistent with later attested Catawba. See Shea (1984:78).

139	*yup* yap 'tree'	Gallatin (1836:337)	Entry is consistent with later attested Catawba. See Shea (1984:253).
140	*eeúp* yap 'wood'	Barton (1798:133)	Entry is consistent with later attested Catawba. See Shea (1984:253).
141	*eeapaúh* ee-up-hah yap-ha 'leaf'	Barton (1798:76)	Entry is consistent with later attested Catawba. See Shea (1984:145).
142	*yunnup-pees* yana-pis 'bark'	Gallatin (1836:338)	Entry is consistent with later attested Catawba. See Shea (1984:253).
143	*surrak* serak 'grass'	Gallatin (1836:338)	Entry is consistent with later attested Catawba. See Shea (1984:118).
144	*yay* 'oak'	Gallatin (1836:339)	Not attested in other lexicons. Later attested Catawba is *watkáheree* is 'oak'. See Rudes (2008b).
145	*eetawa* ičuwe 'pine-tree'	Gallatin (1836:339)	Entry is consistent with later attested Catawba. Note Gallatin did not hear /č/.
146	*weedee-yoyunde-e* wíiduu yądee cow-meat 2s.eat-IMP 'flesh, meat'	Gallatin (1836:339)	Informant clearly indicated here '(you) eat the meat'. Furthermore, this entry clearly shows *wide* in its unsycopated form. See Shea (1984:154).
147	*weedee-youh* wide-yu 'cow-dead'	Barton (1798:116)	Entry is consistent with later attested Catawba. See Shea (1984:154).
148	*chaupee* čápi 'beaver'	Gallatin (1836:340)	Later attested Catawba form can be as follows: *tąsi cap* (Pickens 1957) *tąsi yamuree* (Rudes 2008b)
149	*weedaboy-ah* wita-bo-aʔ cow-shoot-SUB 'deer'	Gallatin (1836:340)	Entry is consistent with later attested Catawba. See Shea (1984:85).
150	*yunnaus* 'bison, buffalo'	Gallatin (1836:340)	Speck has also attested *yedehás* (Speck 1934).
151	*nomeh* *namęʔ* 'bear'	Gallatin (1836:341)	Entry is consistent with later attested Catawba. See Shea (1984:54).

152	*yauntsesoore-ee* tąsi-surie dog-wild 'wolf'	Gallatin (1836:341)	Entry is consistent with later attested Catawba. Note that /t/ was replaced by /y/ according to Miller. See Shea (1984:54).
153	*tauntsee, taunsee* tąsi 'dog'	Barton (1798:83)	Entry is consistent with later attested Catawba. See Shea (1984:54).
154	*dupoyamo-eeha* dapa yámuye 'fox'	Gallatin (1836:342)	Entry is consistent with later attested Catawba. See Speck (1934:story 19).
155	*piup* payą 'squirrel'	Gallatin (1836:342)	Entry is consistent with later attested Catawba. See Shea (1984:212). Note that Miller heard a stop, /p/.
156	*depauksa* dapa-pą-sę something-belly-old 'rabbit'	Gallatin (1836:342)	Entry is consistent with later attested Catawba. See footnote on Speck (1934:9). Note this form is like Oscar Lieber's (1858:14) form.
157	*yah* yaa 'snake'	Gallatin (1836:343)	Lieber (1858:7) also records *daa* for snake. See Shea (1984:205).
158	*koching* kučin 'bird'	Gallatin (1836:343)	Entry is consistent with later attested Catawba. See Shea (1984:55).
159	*watka-eno* witką-hiinų chicken-egg 'egg'	Gallatin (1836:343)	Entry is consistent with later attested Catawba. See Shea (1984:95). Miller missed the nasal quality of the second vowel.
160	*ahhah* ááhaaʔ 'goose'	Gallatin (1836:344)	Entry is consistent with later attested Catawba. See Shea (1984:118).
161	*kosaunsopee* kasąsupi 'duck'	Gallatin (1836:344)	Entry is consistent with later attested Catawba. See Shea (1984:92).
162	*eetoosewee-yachuh* ituse-we yą-čaa dove-?? water-DIM dove-?? creek 'pigeon'	Gallatin (1836:344)	Analysis for this entry is not certain. Also, it is not clear if Miller misheard the suffix *-we*. Speck had previously attested: *itúse bee?* dove immovable 'passenger pigeon' (See Rudes 2008b:49.)

163	*eepahka* ípakee 'partridge'	Gallatin (1836:345)	Later references attested the following: *dapakee* or *dapakeek* dapa ípakee something ??-partridge See Rudes (2008b:18 and 47). It is not clear if the form is *ípa-kee* or *pakee*.
164	*watkunterro* witką taro chicken large 'turkey'	Gallatin (1836:345)	Entry is consistent with later attested Catawba. See Shea (1984:235).
165	*yee* yii 'fish'	Gallatin (1836:345)	Entry is consistent with later attested Catawba. See Shea (1984:66).
166	*saukchuh* sakči 'white'	Gallatin (1836:346)	Later attested Catawba is *takci*, however the archaic construction resembles proto-Siouan *asąą(-ha) (Rankin et al. 2015). If this is truly a cognate, the nasal quality was missed.
167	*houkchuh* hawukči 'black'	Gallatin (1836:346)	Entry shows coalescence of *-awu-* to a back rounded vowel. See Shea (1984:56).
168	*sikechuh* siʔką-či 'red'	Gallatin (1836:346)	Entry is consistent with later attested Catawba. 'Red' has been attested as *siʔką* Rudes (2008b:102) and *skę* Rudes (2008b:103). Miller may have recorded an archaic version of this word; that is, *siʔką* may be short for *siʔkąči*.
169	*yahwe-hah* yą wi-h-aʔ water is.like-3s-SUB 'blue'	Gallatin (1836:347)	'Blue' has been attested as *wayą́kaa* or *wiyą́kaa* Rudes (2008b:122).
170	*sekaweehuh* siką wi-h-ree red-is.like-3s-IND 'yellow'	Gallatin (1836:347)	Entry is consistent with later attested Catawba. This utterance was later attested by Gatschet (1900:534).
171	*wi-unka* wiyąk 'green'	Gallatin (1836:347)	Entry is consistent with later attested Catawba. See Rudes (2008b:122).

172	*pauktehera* paatkę-h-ree is.big-3s-IND 'great, big'	Gallatin (1836:348)	Entry is consistent with later attested Catawba Rudes (2008b:88). However, Miller must have misheard and switched the *k* and the *t* in his notes.
173	*tee-huera* tų-h-ree be.small-STAT-IND 'small, little'	Gallatin (1836:348)	Entry is consistent with later attested Catawba (Rudes 2008b:116).
174	*yahneerochora* yį miiraʔ-ču-ree man be.strong-INTS-IND 'strong'	Gallatin (1836:348)	Entry is consistent with later attested Catawba. See Shea (1984:119).
175	*sebaheh* 'old'	Gallatin (1836:349)	Entry is consistent with later attested Catawba. See Shea (1984:170).
176	*worera* -wariʔ-ree 'young'	Gallatin (1836:349)	Entry is consistent with later attested Catawba, see Voorhis (n.d.:103). This verb requires a pronominal prefix and a suffix.
177	*koonera* kurį-ə-ree good-EP-IND 'good'	Gallatin (1836:349)	Entry is consistent with later attested Catawba. See Shea (1984:116).
178	*imbahow-ara* himbaháaree himba-(h)-aa-ree nice-EP-NE-IND.MODE 'bad'	Gallatin (1836:350)	Entry is consistent with later attested Catawba. See Voorhis (n.d.:41).
179	*koonehara* kurii-h-ree 'handsome'	Gallatin (1836:350)	Entry is consistent with later attested Catawba. See Shea (1984:116).
180	*ehechowehara* eʔ-h-ču wį-h-ree to.dislike-3s-INTS to.resemble-3s-IND 'ugly'	Gallatin (1836:350)	Morphemes are uncertain. Perhaps the informant said 'he resembles he doesn't like'.
181	*yawahrahcha* yį wari-če man dead-PROHIB 'alive, life'	Gallatin (1836:351)	Entry is consistent with later attested Catawba. See Shea (1984:85).

182	*yawahrah hera* yį wariheree man dead+3s+IND 'dead, death'	Gallatin (1836:351)	Entry is consistent with later attested Catawba. See Shea (1984:85).
183	*chehuh cheara* čį ha-čuu-ree cold STAT-INTS-IND 'cold'	Gallatin (1836:351)	Typically, describing words in Catawba inflect for person similar to verbs. However, this word, 'cold' does not appear to inflect for person. After a close observation of Shea (1984:80), it appears likely that the suffix *-ha-* that appears on her entries is that stative suffix. Furthermore, 'I am cold' and 'you are cold' are essentially the same entries; thus, it is likely that this adjective does not inflect for person.
184	*weehuhohechora* wiika ha-čuu-ree hot STAT-INTS-IND 'warm, hot'	Gallatin (1836:351)	Like the word 'cold', the word 'hot' does not appear to inflect for person either. See Shea (1984:131).
185	*derah* de 'I'	Barton (1798:97)	Entry is consistent with later attested Catawba. See Shea (1984:135).
186	*yayah* ye 'thou'	Gallatin (1836:352)	Entry is consistent with later attested Catawba. See Shea (1984:257).
187	*ouwah* owa 'he'	Gallatin (1836:353)	Entry is consistent with later attested Catawba. See Shea (1984:125).
188	*wupchahaora* do 'we'	Gallatin (1836:353)	This may be an archaism. Morphemes are uncertain. See Shea (1984:245).
189	*yayah* wi 'you'	Gallatin (1836:353)	This may have been an error recorded by Miller. The informant clearly attested 'you' (singular) and Miller recorded it as 'you' (plural). See Shea (1984:256).
190	*kera-arrera* owa 'they'	Gallatin (1836:354)	This may be an archaism. Morphemes are uncertain. See Shea (1984:125).

191	*aya* 'this'	Gallatin (1836:354)	Previously unattested
192	*heya* híiyaa 'that'	Gallatin (1836:354)	Entry is consistent with later attested Catawba. See marginalia of Notebook #3 (Siebert 1945a). *Híiyaa* was attested to mean 'yonder, nearly out of sight'.
193	*needem* nitęęp 'all'	Gallatin (1836:355)	Entry is consistent with later attested Catawba. See Shea (1984:45). *Ditęęp* has also been attested.
194	*yahkano* yaką 'many, much'	Gallatin (1836:355)	Entry is consistent with later attested Catawba. See Shea (1984:160).
195	*tou-a-ena* tuwę 'who'	Gallatin (1836:355)	Entry is consistent with later attested Catawba. See Shea (1984:249).
196	No Entry 'near'	Gallatin (1836:356)	
197	*yahpasa* yaap-se day-point 'today'	Gallatin (1836:356)	Entry is consistent with later attested Catawba. See Shea (1984:238).
198	*sodah* sądáa 'yesterday'	Gallatin (1836:356)	Entry is consistent with later attested Catawba. See Shea (1984:256). Recorder missed the nasal quality of the first verb.
199	*yahwah* yawą 'tomorrow'	Gallatin (1836:357)	Entry is consistent with later attested Catawba. See Shea (1984:232).
200	*imbah* 'yes'	Gallatin (1836:357)	Entry is consistent with later attested Catawba. See Shea (1984:256).
201	*wahhow-ara* wahaa-ree no-IND 'no'	Gallatin (1836:357)	Entry is consistent with later attested Catawba. See Shea (1984:173).
202	*noyah* duyą OBJ.+2s.eat 'to eat'	Gallatin (1836:363)	Entry is consistent with later attested Catawba. Often a proclitic *du-* is added to 'to eat'. See Lieber (1858:13). This proclitic appears to be unrelated to the 'hand action' instrumental affix.

203	*korooksa* kuruk-s-aʔ drink-1s-SUB 'to drink'	Gallatin (1836:363)	Entry is consistent with later attested Catawba. See Shea (1984:98).
204	*tereeksera* ciriksiree to.run-1s-IND 'to run'	Gallatin (1836:364)	Entry is consistent with later attested Catawba. See Shea (1984:190). Interesting use of *t* vs. *c*.
205	*barreeda* baa-ree-da dance-IND-?? 'to dance'	Gallatin (1836:364)	Entry is consistent with later attested Catawba. See Shea (1984:90). It is not certain what the affix *-da* means in this index.
206	*koreda* kuuwa-ree-?? to.go-IND-?? 'to go'	Gallatin (1836:364)	Often the /uuwa/ sounds will coalescence into an /o/ sound, as seen in this index. This entry is consistent with later attested Catawba. See Shea (1984:119).
207	*mana* mų-aʔ sing-SUB 'to sing'	Gallatin (1836:365)	Entry is consistent with later attested Catawba. See Shea. Pg. 207.
208	*hemoda* himų-de sleep-IMP 'to sleep'	Gallatin (1836:365)	Entry is consistent with later attested Catawba. See Shea (1984:209).
209	*nedaukwunada* nįda-kąy-de 1s.peak-3s=3s-IMP s/he says to him/her 'to speak'	Gallatin (1836:365)	Entry is consistent with later attested Catawba. See Shea (1984:221).
210	*kawneda* kaani-de see-IMP 'to see'	Gallatin (1836:366)	Entry is consistent with later attested Catawba. See Shea (1984:202).
211	*nummosara* ne-musa-ree 1s-love-IND 'to love'	Gallatin (1836:366)	Entry is consistent with later attested Catawba. See Shea (1984:151).
212	*eekwah* i-kwa-aʔ IDEF-kills-SUB 'to kill'	Gallatin (1836:366)	Entry attests that the indefinite possessor affix can also be the subject of prefixing verbs as well as inalienably possessed objects. Otherwise, entry is consistent with later attested Catawba. See Shea (1984:141).

213	*eewahna* awą̀ʔ-aʔ walk-SUB 'to walk'	Gallatin (1836:367)	It is hard to know if Miller heard a nasal *n* or a vocalic *n*. Swadesh attests to *awą̀-* and McDavid attests to *awaʔ*. *awaʔ-h-ne* 'is he walking?' is also possible. See Shea (1984:241).
214	*nutátie* 'blanket'	Dorsey (n.d.)	This term may be a related term for *núunee?* (See Rudes 2008b:85).
215	*datukó* 'inner bark'	Dorsey (n.d.)	This term may be related to 'in'. See Lieber (1858:11) and see Shea (1984:138).

Appendix B

References for phonology and abbreviations used in this paper:

- Morphemic analysis used in this paper generally follows the phonemes described by Rudes (2003).
- Short vowels are presented as follows: a, e, i, u
- Long vowels are presented as follows: aa, ee, ii, uu
- Nasal vowels are presented as follows: ą, ę, į, ų
- The coalesced vowel is presented as: ɔ
- Siebert (1945a,b) attested that Catawba does not have the vowel /o/ and /oo/. Despite that, these vowels have been used. When a morpheme had previous been attested with /o/ and the authors are not certain what the correct vowel is, the phoneme /o/ is used due to historical reasons.
- Consonants generally follow what was described by Rudes (2003).

The abbreviations used in this paper generally follow Rudes's (2008a) verb morphology paper. The ones used in this paper are described below:

- STAT—Stative Suffix—attested as /-ha-/. It is used with nouns to create a verb phrase. See Index 28 of Appendix A. Mills's consultant was saying 'it is my son' by changing *kuri-na* or 'my son' to *kuri-na ha-ree* or something to the effect of 'it is my son'. We made the stative suffix bold for emphasis.
- IND—Indicative Modal Suffix—Suffix has a phonemic value of *-ree*. Shows statement of fact.
- IDEF—Indefinite possessor—this morpheme was first described by Voorhis's grammar sketch (n.d.). Attested as /i/ and appears as a proclitic to indicate that the possessor of an inalienably possessed noun is not identified.
- INT—Interrogative Modal Suffix—attested as *-ne*. This suffix is used to ask yes or no questions.
- CAUS—Causative suffix—has the phonemic representation as *-č-*. This suffix is used in verb phrases to show the verb was caused by something. Generally, the presence of this suffix increases verb valency.
- SUB—Subordinative Modal Suffix—has a phonemic representation of *aʔ* or ∅. This modal suffix marks a dependent clause.
- NEG—Negative suffix—This suffix has the phonemic representation of *-aa-*.

- GEN—Genuine suffix—This suffix has the phonemic representation of *-ye*. This suffix is affixed to nouns to show the noun is genuine, or true.
- HAB—Habitual suffix—This suffix has the phonemic representation of *-k-*. It is affixed to the verb phrase and means that the verb action is repeated habitually.
- PROX—Proximative Suffix—This suffix has the phonemic representation of *-ra-*. It is affixed to verb phrases and means 'almost' or 'nearly'.
- CONT—Continutive Modal Suffix—This suffix has the phonemic representation of *-hę*. It is affixed to the end of verb phrases to show that the verb 'would' happen.
- IMP—Imperative Modal Suffix—This suffix has the phonemic representation of *-dee*. It is affixed the verbs to show a command.
- INTS—Intensive Suffix—This suffix has the phonemic representation of *-čuu-*. It is affixed to the verb phrase. It shows intensification of the verb.
- PROHIB—Prohibitive Modal Suffix—This suffix has the phonemic representation of *-če*. It is affixed to the verb phrase. It shows prohibition of the verb.

Considerations for the continuation of a Kansa corpus*

Justin T. McBride
Northeastern State University

Abstract: More than a decade ago, the Kaw Nation Language Department began a grant-funded project to compile the known Kansa texts for the purpose of creating a graded reader book. This project resulted in, among other things, a unique XML-based electronic corpus of the language. While valuable in terms of the initial project, a closer look at the corpus reveals its numerous practical problems, including general incompleteness, an overly specific purpose, limitations of use and potential users, and compatibility with modern computing tools. Nevertheless, it could be expanded and modified to serve as a much more functional Kansa-English bilingual aligned corpus usable by both Kaw Nation citizens and language scholars outside of the tribe. Various features of such an expanded corpus and its possible development are considered.

Keywords: Dhegiha Siouan, bilingual aligned corpus, XML, planning

1. Background

The Kansa language, known also as either Kanza or Kaw among members of the Kaw Nation, belongs to the Dhegiha branch of Mississippi Valley Siouan, where it occupies a very close relationship to Quapaw, Omaha, Ponca, and especially Osage. There are presently no L1 speakers of Kansa, but the Kaw Nation has maintained an active and successful language revitalization program for more than two decades. The original project described below was one such activity from this program.

2. Original corpus project

As reported in McBride (2009a), the Kaw Nation Language Department received a grant from the Administration for Native Americans in 2008 to develop a graded reader document with accompanying audio files and making use of texts in the language. Less than three dozen in number and collected by such field researchers as Dorsey (c. 1880) and Rankin (c. 1974-2011), these few texts represent—with the exception of two prayers—the whole of extended monologic discourse recorded from L1 Kansa-speaking consultants and span several genres; no dialogic texts for Kansa are known to exist. The development of the reader package for Kansa language learners was a large

*I would like to extend my sincere thanks to the attendees of SCLC 39 for making numerous useful suggestions about the project described herein and especially to the editorial reader who provided thorough and thoughtful feedback on my original manuscript. All remaining mistakes are entirely my own.

project that involved a great deal of planning and yielded several useful products. The choices made during this project from more than a decade ago still remain relevant today.

2.1. Purpose and products

In the initial planning for the grant, it became obvious that some of the available texts had greater pedagogical value than others. As such, planners at the Language Department decided that a planned reader and supplementary audio CD would offer only a subset of the texts. Moreover, given the rare opportunity to work so closely with the texts, the three linguists associated with the project—then Language Director Justin McBride, then Language Coordinator Linda Cumberland, and the late Language Consultant Robert Rankin—wanted to analyze them as fully as possible. Thus, the three linguists began a systematic morphemic parse of the body of Kansa texts in order to provide the team's Community Advisory Group of tribal citizens—a group whose role was to offer guidance and support solely for this project before disbanding—sufficient knowledge to make an informed decision on which texts to include in the reader while simultaneously generating comprehensive interlinear gloss material for all of the texts.

While the initial plans involved only the assemblage of various electronic files and an audio CD as grant deliverables, a decision was also made around this time to publish the material developed in the project as two separate print volumes. One of these two print works (i.e., McBride & Cumberland 2009) would be the more inclusive of the two and would contain a detailed morphosyntactic analysis and a corresponding comprehensive morpheme glossary for all the texts. This document would appeal to a somewhat more scholarly audience and would be archived for future reference. Owing to the unexpected cancellation of a mandatory grantee meeting at around the same time, grant funds sufficient to print this volume as a short run of only a few dozen copies became available suddenly. The team jumped on this opportunity, and several copies of the first printed volume were produced and archived while others were given away to team members and tribal administrators; still others were donated to select libraries. The second print volume (i.e., McBride & Cumberland 2010) would be the actual graded reader featuring the smaller set of texts; it would be produced following the conclusion of the grant project to ensure a maximum of federally funded effort went into its creation. Thus, the rest of the initial grant project was spent developing this volume, which included illustrations, grammatical explanations, exercises, a glossary, an accompanying CD of audio recordings, and various other learner resources. Additional funding was then secured from the Endangered Languages Fund to print 500 copies of this document for sale to interested learners; many copies of the reader volume remain for sale through the Kaw Nation at the time of this writing.

The two print volumes themselves, while of potential interest to any number of individuals, were in fact merely products of the background system used to compile the text corpus and its analysis. The selection and use of this system are especially germane to the discussion at hand.

2.2. Initial corpus considerations

Project linguists at the Language Department were aware of the fact that detailed work on the Kansa texts would require much effort insofar as the extant analysis was both inconsistent and intermittent. Consider, for instance, that several of the texts collected by Rankin (c. 1970-1979) and especially Dorsey (c. 1880) had been fairly well analyzed, but those analyses were not always theoretically

congruent with each other; Dorsey's (c. 1880) Kansa work, less complete than his work elsewhere in Dhegiha (cf. Dorsey 1890), did not always seem to exhibit great consistency even within itself. Additionally, some texts had never been analyzed. Any analysis associated with them would need to be brought into alignment with both Dorsey and Rankin. Plus, while Rankin's (2008) lexical file for Kansa did manage to feature many items from Dorsey, it was not an exhaustive reconsideration of Dorsey's (c. 1880) Kansa work and did not feature items from such other collectors of texts as Spencer (1908) or Morehouse (c. 1908). In short, a much more uniform processing of all texts was needed.

At the outset of the project, two popular software systems offered the analytical parsing and interlinearization functionality needed to complete the required text processing. One of these was SIL's "The Linguist's Toolbox" (Toolbox), and the other was the American Indian Studies Research Institute (AISRI) at Indiana University's suite of programs including Annotated Text Processor and Indiana Dictionary Database (ATP and IDD). Both of these systems had various pros and cons. For example, while the SIL software was free and Unicode-compliant and offered semi-automatic interlinearization through its seamless text and dictionary integration, it was essentially unsupported by the developers, was difficult to configure for complicated functions, and made use of unintuitive procedural workarounds for routine phonological processes. Meanwhile, although the Kaw Nation staff linguists were already familiar with ATP and IDD and had an established partnership relationship with AISRI, which both developed and supported these programs, ATP and IDD were not so seamlessly integrated, were not Unicode-compliant, were similarly difficult to configure, and were also known to have some nagging performance issues.

In considering the advantages and disadvantages of these two software systems, the team arrived at a simple list of program requirements necessary for simultaneously processing the Kansa texts and compiling an electronic corpus for the language. This list would drive our decision as to which system we would choose to complete the project. The chosen system would need to meet the following feature criteria:

- *Free*—the grant budget did not allow for the purchase of specialized software intended exclusively to help with the analysis of texts;
- *Amply supported*—the project's narrow timeframe was not conducive to attempting overly problematic solutions, and the availability of rapid, high quality support to address possible obstacles was considered critical to project success;
- *Unicode-compliant*—regardless of the fonts chosen for spelling Kansa words, the practical orthography for Kansa makes use of various accented vowel characters, <á, à, é, è, í, ì, ó, ò, ú, ù> (representing primary and secondary stress, respectively), and one high-frequency but potentially problematic character, <ⁿ>, whose use in marking vowel nasalization could not be avoided;
- *Cross-platform*—because the Kaw Nation offices, where the work would be completed, were equipped exclusively with IBM-compatible computers while many Kaw tribal citizens preferred Macintosh computers, the solution had to be compatible with both systems;
- *Self-contained*—the solution could not be part of a larger system that end users would also have to obtain and then learn to use; and

- *Small*—the solution would have to create a text corpus that could be easily shared between team members via USB flash drive or e-mail.

2.3. Solution

Rather than choose a pre-packaged application that only met only some subset of these requirements, the team ultimately opted for a self-designed XML-based solution that could be developed entirely in-house and still achieve the desired functionality. Extensible Markup Language, or XML, is a coding system that allows pre-existing text (i.e., the content) to co-exist with embedded computer processing instructions (i.e., the code), the latter of which can be used to manipulate the former. Think, for instance, of how a webpage coded in the related Hypertext Markup Language, or HTML, works to display the content of the page in a particular manner when opened in a browser window. In HTML, the code to make a passage of content text appear italicized begins with an opening "tag" placed before the text in question and ends with a closing tag placed afterward, as in 'Hello, friends', which yields '*Hello, friends*' in italics. XML works similarly, but it is different in a very important way: XML is generic and lacks established tag codes for the embedded processing instructions. Rather, the codes and the instructions they represent are largely left up to the coder to define by way of a supplementary 'stylesheet' document; this makes XML potentially much more powerful than HTML, which is limited as to what it can do by its range of pre-existing codes. Given XML's flexible nature, an XML-based approach to developing the Kansa corpus would give the team the freedom it needed to create the corpus it desired directly from word processed versions of the texts. XML, moreover, is free, supported by numerous online developer communities, Unicode-compliant, cross-platform, and self-contained and results in comparatively small files that can be easily shared.

After making the decision to go with an XML-based solution, the team began developing the necessary files to realize this plan. Two XML documents had to be developed along with their two corresponding stylesheet documents. One of these two XML documents would include the texts themselves serving as content. Embedded within, the code for this document would also include the line-by-line and morpheme-by-morpheme parse and any notes or other such supplementary material for each text, such as the consultant's name, the date of collection, reference to any illustrations, etc. To populate the ultimate interlinearization of the texts and to generate a constantly updated glossary, a second XML document would in turn include the morphemic units themselves along with a gloss, lexical and semantic classes, and a numeric code for each. To relate the two documents, the numeric codes from the glossary document were referred to in the text document rather than the morphemes themselves. That way, a change in the morpheme document to any single entry would cascade throughout the interlinear analysis in the text document.

The two associated stylesheet documents were written to generate the desired output materials, namely, 1.) a body of parsed and annotated texts and 2.) a glossary of all the morphemes appearing in these texts; the latter document also drew an example sentence for each entry from the former document and listed the location of the sentence within the corpus. Additionally, 10 illustration graphics resided in the same folder as the two XML documents and their two stylesheets. Even with these graphics, the entire folder was under 1.5 MB and could be easily transferred from computer to computer using almost any sharable media—even the then increasingly rare 3.5-inch double-density floppy disk. The only software needed for any user to access the material in a usable format was a free web browser, e.g., Explorer, Firefox, or Safari, capable of compiling XML code,

which, even at the time, was a standard feature on up-to-date browsers.

Of course, interlinear text is not generally what is meant by the term 'corpus,' which refers only to a body of texts. Nevertheless, a simple modification of the stylesheet—the inclusion of two small tags to demote stylesheet code from instruction to comment—could be used to generate text that could be used with any free concordance software such as AntConc or MonoConc. From this, routine corpus work could be done using the admittedly small body of Kansa texts.

2.3.1. Corpus contents

Using the system described above, the following extended monologic texts were converted for use in the electronic Kansa corpus:

- eight myths (for lack of a better word), thirteen personal histories, and three items of personal correspondence from Dorsey (c. 1880);
- one song from Spencer (1908);
- one transcribed speech from Morehouse (c. 1908); and
- five myths transcribed from Rankin (c. 1970-1979).

The only other pieces of extended monologic text from L1 Kansa speakers consist of two prayers that have never been adequately parsed and are generally assumed to be of a sensitive religious nature. These prayers were omitted from the electronic corpus, as were all other known Kansa materials that did not contain extended monologic texts, such as word lists or even sentence-length elicitation responses.

3. Practical evaluation

Several observations can be made about the use of this XML-based solution for the problem of compiling and analyzing Kansa texts and generating pedagogical materials from them. On the one hand, it worked! That is to say, the two planned volumes were successfully produced using the newly developed corpus materials. Additionally, the corpus is still available for additional computer-assisted study of the Kansa language through the viewable front-end output of the XML files, which can be manipulated in various ways via the back-end interface to reveal language data in new and thought-provoking ways. On the other hand, there are still many issues that were never dealt with, some that did not even occur to the planning team at the time of the development of the XML corpus.

3.1. Problems

The most obvious problem involves the completeness of the corpus. The original set of compiled and parsed texts included only extended monologs of lengths greater than a single clause which were collected by Dorsey (c. 1880) and Rankin (c. 1974-2011) plus two others (i.e., excerpts from Spencer 1908 and Morehouse c. 1908). This is far from the entirety of sentential material collected from Kansa speakers. The largest source of additional material is Rankin, whose (c.

1970-1979) field notebooks alone conservatively contain over five times the amount of material in the corpus at present; compare the approximately 4,200 lines of known notebook material to the approximately 800 lines of compiled corpus texts. These notebooks document elicitation sessions with three separate speaker consultants, both male and female, and span the better part of a decade. Not all the material is appropriate for rigorous corpus-based analysis insofar as it is often no more responses to requests for single Kansa words or phrases. The speakers often struggle to recall such words, resulting in many false starts and obvious mistakes. Worse still, the clausal material that does appear is merely a response to an elicitation and, in terms of broader discourse-level considerations, unconnected to what comes before or after. Nevertheless, sentence-level material is available for all three of these L1 Kansa-speaking consultants.

Moreover, while working with his primary consultant, Maude Rowe, Rankin stopped collecting material in notebooks and shifted over to eliciting responses straight from photocopies of Dorsey's original Kansa dictionary slip files, handwriting Rowe's responses directly on these copies (Rankin c. 1974). While Rankin did allow the Kaw Nation to make further copies from his annotated Dorsey slip files, these have not been systematically examined to collect sentence-level material for inclusion in any digital document. A smattering of additional material may also be available, for example, in the extensive Bourassa (1843)[1] and Morehouse (c. 1908) collections for the Kansa language. At present, no known clausal material from these collections remains unanalyzed, but more research is needed to be sure.

Another problem arises from the purpose behind the corpus. Specifically, the XML solution was developed for very particular goals involving text-based language pedagogy. This was its primary purpose, and general language scholarship was only a happy consequence. Clearly, the choices made with the pedagogical goal in mind affected the design of the system, which in turn creates obstacles that must be dealt with for more routine corpus work. For instance, it has already been mentioned that modifications must first be made to the stylesheets to generate output usable by standard, third-party concordancing software, which must also be obtained elsewhere. The inconvenience of these first steps makes even a simple keyword-in-context search a tedious process.

Given the built-in purpose, even the range of potential uses is somewhat limited. One logical use of the XML materials, for example, would be the subsequent development of a stand-alone multilingual aligned corpus. This category of corpus includes such corpora as Compara, which is a bilingual Portuguese-English corpus that can be queried in various ways in either language (cf. Frankenberg-Garcia & Santos 2003), or *MulTed*, a proposed multilingual corpus composed of TED talk titles and subtitles (Zeroual & Lakhouaja in press). At present, configuring the Kansa materials in this way would be very time-consuming, mostly because of part of speech tagging

[1]There is no convenient means of citing or even referring accurately to the Bourassa materials. Consider the following personal communication from Ives Goddard from August 11, 2008, alerting Robert Rankin of their existence: "The Cullman library in the Smithsonian Natural History Museum has acquired a ms. with vocabularies of Potawatomi, Ottawa, and 'Kaw' which is annotated by Wilberforce Eames but apparently copied by someone else from original mss. of Joseph N. Bourassa. (A ms. related in some way is in the Pequot library in Conn.) The ink is faded and often hard to make out even with the naked eye, but much is readable and interesting. The 'Kaw Dictionary' (on pp. 163-183) is probably copied from the one listed for Bourassa by Pilling (then in the possession of John B. Dunbar), and at least one of the Potawatomi sections is presumably a copy of the Potawatomi vocabulary that Pilling also gives as Dunbar's. The Ottawa is sandwiched between two Pot. sections in the copy we have, and as it is not labeled as such the exemplar may have gone unidentified." After receiving this message, Language Department staff obtained a photocopy of the 'Kaw Dictionary' excerpt mentioned above directly from Goddard at the Smithsonian Institution while on a work-related trip to Washington, DC.

for the Kansa, which would have to be done manually; the current tagging is not strictly at the word-level. The parsing of the Kansa texts is currently morpheme-by-morpheme, meaning that all morphologically complex words in the texts would have to be coded for word-level lexical class. This would require a new level of interlinear analysis that would have to be developed for every word of every text. Note, by the way, that the corresponding English tagging would not be as difficult—it could be done automatically through a part of speech tagging program such as TagAnt or TreeTagger—but an additional line of analysis would have to be added to accommodate the tags for English just as with Kansa. This is to say nothing, of course, of the theoretical concerns about lexical class in Siouan as a whole. For instance, some may argue that Kansa has no adjective class, but only stative verbs, while others may disagree; it is impossible to expend the effort on tagging or expanding the available analysis without opening numerous of such cans of worms, and the results could potentially decrease the potential number of users.

The current XML-based solution already has a very, very small number of users. While the McBride & Cumberland (2009) volume includes the current analysis resulting from the corpus compilation, less than 40 copies were published, and many of these copies reside in archives or are owned by individuals who may lack the necessary experience with interlinear analysis to make ample use of it. This means that the work done for the corpus project is mostly left up to users of the XML-based source files. Given that the coding structure is unique, that there is no convenient query interface (corpus queries can be approximated by simple search functions in word processing based off of the numeric codes associated with individual morphemes), and that manipulation of the source files requires learning XML, any use of the source files outside of their primary purpose involves a steep learning curve. There may only be a handful of people comfortable using these files for anything despite the potential value the files may possess.

Finally, the XML files are no longer easily viewable on browsers. While the display of lengthy local XML code by way of an associated local stylesheet is still possible on some browsers (e.g., Edge and the now obsolete Internet Explorer), it is rare enough that extra steps must taken to do so on some browsers (e.g., Chrome requires Document Type Definitions to compile the files and recommends use of its XML Viewer extension), and it is simply not possible on others (e.g., Firefox).

4. Present considerations

With so much material that could be converted for use with the Kansa corpus documents, and with so many serious complications associated with the current corpus, it would seem that the project stands at a crossroads. In order to decide how to proceed, several questions must be answered.

4.1. Who would use such a corpus and to what ends?

Clearly, the project should benefit Kaw Nation citizens first and foremost. The work, after all, began as a grant-funded tribal project to produce materials for tribal learners and made extensive use of resources furnished by the tribe to facilitate its completion. Any derived products would also need to be geared toward primary use by Kaw Nation citizens, presumably as an interactive archive of knowledge relating to their heritage language. The final product should, therefore, be targeted to an audience composed mostly of non-specialists in language study without relying heavily on

theoretical terminology, niche technology, or impractical functionality that cannot readily advance language learning. As before, an advisory group composed of tribal citizens could be assembled to provide guidance on how work on the project should progress, all the while keeping the tribe's best interests in mind.

Secondarily, the design should permit Siouan scholars, language professionals, and others with a vested interest in the promotion of understanding of Kansa and its related languages to achieve their goals. For example, while simple corpus tasks such as keyword searches or even line-by-line navigation through texts should be obvious to non-specialist users, the functionality must be robust enough to allow for much more complex use of the language data in ways that may not immediately occur to such users. The interface design must also not appear to hide such functionality from non-specialists. Moreover, Dorsey's (c. 1880) and Rankin's (c. 1974-2011) Kansa materials, which are of special interest to scholars for their relative regularity and overall trustworthiness, should be as comprehensive as possible within the data; materials from as many others as can be managed should also be included.

Another feature that would be potentially valuable to scholars would be the ability to toggle between the practical spellings of Kansa and Siouanist phonemic transcriptions; the former are actually derived from the latter, but they can obscure more complex phonological goings-on within the language, especially with regard to phenomena that may be of cross-linguistic interest. For example, the practical Kansa <p, t, k> characters correspond to the Dhegiha 'tense' stop series that is realized as /pp, tt, kk/ in Kansa, Omaha, Ponca, and Quapaw, and as /hp, ht, hk/ in Osage—not as plain /p, t, k/ in Osage and Quapaw, which correspond to /b, d, g/ in Kansa, Omaha, and Ponca. Consider, for instance, the word for 'similar, alike' in Osage /kǫzékǫ/ and Kansa /góze égo/, practical <góze égo>, where the plain stops surface in both languages; the plain Osage stop /k/ corresponds to Kansa phonemic /g/ and practical <g>. But, consider 'teaching, religious devotion' in Osage /hkihkǫ́ze/ and Kansa /kkikkáze/, practical <kiká^{n}ze>, where the tense stops surface in both languages; the tense Osage stop /hk/ corresponds to Kansa phonemic /kk/ yet practical /k/. On a similar note, <aá^{n}> in the Kansa practical orthography corresponds to a long, nasal vowel with falling pitch, which Rankin tends to represent as /ą̀ą̀/ in his (1974-1975) notebooks. As such, the practical spelling of the name of the tribe and the language, <Kaá^{n}ze>, which tribal citizens have come to accept, corresponds to the more familiar Siouanist transcription /kkąą̀ze/, which tribal citizens may not even recognize as the same word. While these orthographic concerns may appear minor, the division between practical and technical spellings for Kansa is an important one, and it has been the subject of intense internal debate and planning (McBride 2009b).

4.2. What form should it take?

This single question is in many ways far more nebulous than the first. One the one hand, a list of desirable design features should be easy to come by. On the other hand, some of the most fundamental considerations that would be helpful for such a design wish list have never been dealt with. For example, it is not clear at present if converting the remaining data to the current corpus format, which could then be modified wholesale to develop the desired corpus tool, is preferable to starting essentially from scratch, saving only those parts of the current corpus needed for the optimal end state, whatever that may be. At any rate, there are still some things that must be accomplished.

4.2.1. Bilingual alignment—for a start

The end product must involve a multi-use bilingual corpus interface that involves multiple levels of analysis. At the most basic level, parallel text must be available for practical and/or phonemic Kansa on one hand and a corresponding English gloss on the other hand. An interlinear parse, perhaps featuring a user-defined depth, should also be immediately available. A close phonetic transcription option may also be desirable, but it would not be available for any of the texts save for those collected by Rankin, whose (c. 1970-1979) audio recordings of the elicitation sessions survive.

4.2.2. Robust user functionality

The version of the product finally released for public use must include various searching, sorting, multimedia, etc. tools usable as necessary by its two main groups of end users, Kaw tribal citizens and language scholars. One potentially very valuable search tool that would be of interest to scholars is the ability to search for two items in the same sentence. In Dhegiha, determiners associated with the subject often—but not always—take the same form as auxiliaries in the predicate; being able to search for both—either by lemma or by part of speech tag—could help to clarify the reasons for this. Such functionality could be achieved by allowing additional n-gram searches on initial search results. Frequency-based sorting of concordance results would also be very helpful, as would the ability to align audio recordings, where available, to Kansa sentences. Beyond this, routine corpus tools, wherever possible, should be included.

4.2.3. Additional information about texts

Elicitation conditions must be made clear. Highly contextualized and discursively complex material from an extended monologic text may be found in the data right alongside single-sentence responses to very simple elicitation requests. While each is valuable in its own right, an instance of the second lacks the cohesion and coherence of a single sentence from the first; comparing the two is a proverbial apples-to-oranges scenario. Meta-data on the corresponding texts, speaker consultants, collection dates and times, and other such general reference data, must be recoverable from any single line of text or even individual words from it. Recovery of second-order data as this can provide users the analytical context needed to judge how best to interpret the primary data.

4.3. How should it be delivered?

This question flows from the last one, but specifically frames the consideration in terms of which technological solution will maximize ease of the use while also minimizing conversion obstacles and possible errors. On the one hand, one obvious solution would be a web-based platform making use of SQL databases for storing the data and an interface whose functionality would be enabled by PHP scripts calling on the data. This sort of system is far more common these days than the XML documents and related stylesheets found in the original corpus. On the other hand, migrating the entirety of the current system to a format that is not at all similar would be tedious and time-consuming. Moreover, a stand-alone app usable on mobile devices may be even more popular for end users. Given the very small file size of the current corpus, such a solution may be particularly efficient. At any rate, some of these questions could be deferred until such time as another

advisory group or other such ad hoc tribal panel could be assembled to provide culturally sensitive suggestions and guidance.

5. Conclusion

In this paper, I have attempted to demonstrate some of the considerations associated with the creation and possible continuation of an electronic corpus of Kansa texts.

5.0.1. Summary

The Kaw Nation Language Department developed an electronic Kansa corpus more than a decade ago for the purpose of creating pedagogical materials for Kaw tribal citizens. This original corpus is still operational, but it suffers from several deficiencies. While representing nearly the whole of extended monologic discourse in Kansa, it is noticeably incomplete, having been developed from many different sources of many different kinds, and its purpose and resultant structure at present are too narrow for a wide range of uses; as such there are few, if any present-day users. It is also difficult to access today given how technology has advanced since its initial development. Nevertheless, given both the incompleteness and the trove of additional material that could be adapted for use with the corpus, now is an ideal time to review its condition with an eye toward its possible use in the future.

Balancing the interests of the two primary stakeholders while expanding on the original project is obviously a key concern here. On the one hand, the corpus should continue to be used for the primary benefit tribal members seeking to help revitalize their heritage language or who may simply wish to learn more about what their fellow tribespeople had to say *in their own words*. Part of this tribe-centered purpose would also directly benefit the Language Department who are constantly looking for new and exciting—but ultimately simple—ways to generate meaningful pedagogical content for their students. On the other hand, it could also be used by scholars and language teachers and learners from outside of the Kaw Nation who, although somewhat secondary to the main purpose, may find the prospect of relatively unfettered access to a body of lesser known Dhegiha Siouan texts appealing and who may use the corpus to advance particular theoretical or practical goals.

Assessing the situation from these vantage points yields a veritable wish-list of features stemming from the initial development of the corpus, yes, but also poised to govern all aspects of its expansion. The resultant system should have the following features:

- *Free*—The corpus should cost no additional money to develop, maintain, and access (implicit here is that the responsibility for development and maintenance should remain with the Kaw Nation Language Department and its affiliates as directed by a panel of concerned tribal citizens—just as it was at the beginning of the project);
- *Supported*—The technology driving the corpus must enjoy ample development support;
- *Unicode-compliant*—Representation of the language, especially if fields are added to allow for technical phonemic transcription, requires an expanded range of characters in any typeface used for the project (implicit here is that the typeface should be free for the end user to access);

- *Cross-platform*—Just as before, the corpus should be usable on many different platforms (implicit here is that mobile devices, which were not a concern at the time of the initial development, are likely to be a major driver of choices in the expansion of the corpus);
- *Self-contained*—Now more the ever, the corpus should not rely on external resources to use (implicit here is that, if downloadable, development of the expanded corpus should include some sort of installation tool to ensure that the various components are in place and working properly, in terms of both what is available at the time of initial download and also what may be added later as a result of updates);
- *Small*—The corpus should have a small digital footprint (implicit here is that whatever delivery technology is used for expanding the corpus does not unjustifiably add to the overall size of the accessible content);
- *Primarily Kaw-oriented*—The expanded corpus must benefit Kaw Nation citizens primarily (implicit here is that Kaw Nation Language Department staff members are likely to be its most prolific users, but those without technical skills in language description or teaching will ultimately benefit from its expansion the most);
- *Secondarily academic*—The corpus should benefit other Kansa language scholars, teachers, and learners, albeit secondarily (implicit here is the assumption that Siouanists are likely to be among its users, and all onboard functionality should at least be congruent with popular theoretical and practical understandings of Siouan languages);
- *Reflective of different attitudes regarding spelling*—The actual content of the expanded corpus potentially alienates prospective users unless both practical and technical spellings are recoverable (implicit here is that a toggling function and the underlying mechanism for ensuring its effective use must be built-in to either the content or code or both);
- *Bilingual*—At a minimum, the corpus should feature bilingual alignment between Kansa and English sentences where appropriate, but additional functionality may extend to finer-grained levels of analysis on either side (implicit here is the belief—which may well be unfounded—that discourse in one language may align sentence-by-sentence with discourse in another language, and perhaps even at lower levels);
- *Feature-packed*—The corpus must provide robust user-functionality (implicit here is that a survey of potential Kaw and non-Kaw users may need to be conducted to be sure of which features will be the most effective for achieving specific goals);
- *Metadata-packed*—The corpus must provide additional information about the texts sufficient for understanding the place of a single sentence within the corpus—either as a stand-alone item or as an element of a larger discourse—and the circumstances of its utterance wherever known (implicit here is a more or less complete understanding of these considerations, not to mention an efficient means of encoding that understanding into the system); and
- *Convenient*—Along the same lines as self-contained and small above, the corpus must be accessible in a convenient and popular digital format (implicit here is that technologies are known to change quickly, and that subsequent development take a long-view with respect to future use of the corpus).

To ensure that all of these conditions are met in the final product and that the corpus retains its usefulness for Kaw citizens, it is also essential that a panel of such tribal stakeholders provide some degree of oversight on the project, exactly as was done before.

5.0.2. Looking forward

Until a new advisory panel can convene to provide definitive guidance on how to proceed under the following conditions, entering clausal material from the Rankin (1974-1975) notebooks into the current corpus scheme should not be difficult. What is more, the available morphemic data can already be used to populate much of this new material without having to add new morpheme entries. Completing this task for the available notebooks, therefore, should prove a satisfying preliminary step before work could commence on locating other such sentential data in the Rankin (c. 1974) dictionary materials or the materials collected by other researchers; for instance, looking for heretofore unknown clausal material in the Bourassa (1843) or Morehouse (c. 1908) materials would be an excellent idea. Provided that additional texts for inclusion in the corpus cannot be found among these physical sources of written Kansa, the extensive audio recordings Rankin's (c. 1970-1979) dictionary elicitation sessions could additionally be re-transcribed in an effort to locate clausal material. Whatever the case may be, it is hoped that, by the time the notebook material has at last been entered, a more permanent form for the Kansa corpus will have presented itself.

References

Bourassa, Joseph N. 1843. A vocabulary of the Po-da-wahd-mih language [manuscript]: With illustrative sentences and a translation of the first three chapters of the Gospel of Matthew, followed by a vocabulary of the Kaw language. Ms. PM2191 .Z5B68 1843a. Washington, DC: Joseph F. Culman 3rd Library of Natural History, Smithsonian Institution.

Dorsey, James Owen. 1890. The Ȼegiha language. *Contributions to North American Ethnology* 6. 1–794.

Dorsey, James Owen. c. 1880. Kansa texts. James O. Dorsey papers. Suitland, MD: National Anthropological Archives, Smithsonian Institution, MS 4800, Box 33, Items 245-247.

Frankenberg-Garcia, Ana & Diana Santos. 2003. Introducing COMPARA: The Portuguese-English parallel corpus. In Federico Zanettin, Silvia Bernardini & Dominic Stewart (eds.), *Corpora in translator education*, 71–87. Manchester, UK: St. Jerome.

McBride, Justin T. 2009a. Compilation of Kansa texts: Overcoming challenges in the drafting of a graded reader. Paper presented at the 29th Annual Siouan and Caddoan Langauges Conference. University of Nebraska-Lincoln, 12 June. Lincoln, NE.

McBride, Justin T. 2009b. Orthography and ideology: Examining the development of Kaw writing. In Daisy Rosenblum & Carrie Meeker (eds.), *Proceedings from the 12th annual workshop on American Indigenous Languages*, vol. 20 Santa Barbara Papers in Linguistics, 30–45. Santa Barbara, CA: University of Santa Barbara Department of Linguistics.

McBride, Justin T. & Linda A. Cumberland (eds.). 2009. *Compiled Kanza texts*. Kaw City, OK: Kaw Nation.

McBride, Justin T. & Linda A. Cumberland (eds.). 2010. *Kaáⁿze wéyaje (Kanza reader): Teaching Kanza literacy through historical texts*. Kaw City, OK: Kaw Nation.

Morehouse, George Pierson. c. 1908. eorge pierson morehouse papers. Manuscripts Collection 453. Topeka, KS: Kansas State Historical Society.

Rankin, Robert L. 1974-1975. Series 2: Kaw (Kansa, Kanza); 2.1: Field notebooks. Robert Rankin papers. NAA.2014-16, Box 2. Suitland, MD: National Anthropological Archives, Smithsonian Institution.

Rankin, Robert L. 2008. Kansa-English lexical file. Ms. Lawrence, KS: University of Kansas.

Rankin, Robert L. c. 1970-1979. Series 2: Kaw (Kansa, Kanza); 2.3: Sound recordings. Robert Rankin papers. NAA.2014-16, Boxes 51-54. Suitland, MD: National Anthropological Archives, Smithsonian Institution.

Rankin, Robert L. c. 1974. Series 2: Kaw (Kansa, Kanza); Kansa-English dictionary vocabulary slip files. Robert Rankin papers. NAA.2014-16, Boxes 42-45. Suitland, MD: National Anthropological Archives, Smithsonian Institution.

Rankin, Robert L. c. 1974-2011. Series 2: Kaw (Kansa, Kanza); 2.2: Files. Robert Rankin papers. NAA.2014-16, Boxes 3-4. Suitland, MD: National Anthropological Archives, Smithsonian Institution.

Spencer, Joab. 1908. The Kaw or Kansas Indians: Their customs, manners, and folk-lore. *Transactions of the Kansas State Historical Society* 10. 373–382.

Zeroual, Imad & Abdelhak Lakhouaja. in press. MulTed: A multilingual aligned and tagged parallel corpus. *Applied Computing and Informatics* .

Functions of the prefix *wa-* in Umoⁿhoⁿ*

Julie Marsault
Université Sorbonne Nouvelle – Paris 3

Abstract: The verbal prefix *wa-* is well attested in all branches of the Siouan languages. It has several functions, especially in Umoⁿhoⁿ, and this creates difficulties in analyzing it, and in teaching how it works in a clear way. I propose an analysis dividing wa- into 3 different functions: 3rd person plural object marker; antipassive marker; and nominalizer. Most importantly, I identify two types of antipassive constructions with *wa-*, according to its referentiality. While antipassives with generic objects are ambiguous with 3rd person plural object marker, antipassive with referring objects do not lend themselves to such ambiguity, and are rather close to the nominalizing function. This creates difficulties in analyzing wa- synchronically, and also casts doubt on the possible diachronic origin of the antipassive.

Keywords: antipassive, object marking, object referentiality, nominalizer

1. Introduction

The prefix *wa-* is attested with similar functions in all branches of Siouan languages and has already been the subject of studies in a comparative perspective (Boyle 2009a,b). It has been reconstructed in Proto-Siouan as an absolutive marker **wa-* or **waa-* (Carter et al. 2006:928). In descriptive grammars or sketches of particular Siouan languages, *wa-* and its cognates is often attributed an object-removing function, differently called "valency-decreasing" (Boyle 2009a), "detransitivizing" (Hartmann 2015:1270), "absolutive" (Carter et al. 2006:928), "indefinite object marker" (Ullrich 2008:735) or "unspecified argument" (Kasak 2019:231). It is also commonly attested on nouns for objects or agents, with the authors either integrating the examples of nouns as instances of object demotion, or treating it as a distinct function of nominalizer. Some examples of these functions are presented in (1) to (4). In all this paper, I underline the objects of transitive verbs, both in the Siouan language version (when they are expressed by NPs) and in the corresponding English translation. The *x* stands for the object position in unmarked forms (the object position *x* is not generally explicitly mentioned by the authors cited here, but it is understood from their descriptions that the verbs are transitive).

(1) Crow

a.	*íkaa*	'to see *x*'	
	baa*íkaa*	'to see things, to have a vision'	(Graczyk 1991:233)
b.	*dichíichi*	'to boil *x*'	
	baa*lichíituu*	'boiled meat (with plural marker *-uu*)'	(Graczyk 2007:48)

*I am thankful to both my dissertation advisors, Pollet Samvelian and Guillaume Jacques, for their reviews and comments on previous versions of this work. I am also thankful to John Boyle and B. Gordon for the discussions we had about the values of *wa-* in Siouan and in Umoⁿhoⁿ.

(2) Hidatsa (Boyle 2009b:5)
báca 'to string *x*'
maa*báca* 'to string something', 'beads'

(3) Osage (Quintero 2004:§4.1)
čséðe 'to doctor *x*'
wa*čséðe* 'to doctor folks'

(4) Ho-Chunk (Hartmann 2015:1270-1271)
ruuk 'to eat *x*'
wa*rúc* 'to eat something'

Wa- is commonly translated into English as 'things' or 'something' or 'people', or not translated at all. In some cases, *wa-* refers specifically to one kind of object, as in examples (1b) and (2).

Additionally, in the Dhegiha languages and Hochunk-Chiwere languages (Mississipi Valley Siouan), *wa-* serves as a marker of 3rd person plural objects (O.3PL). To the best of my knowledge, the historical source of this function and its possible link with the valency-decreasing *wa-* is unknown.[1]

The purpose of this paper is to explore the distinct functions of the prefix *wa-* in Umoⁿhoⁿ, a Siouan language of the Dhegiha group (Mississipi Valley Siouan), in order ultimately to better understand its historical source. I will particularly focus on its antipassive functions.

This study is based, on the one hand, on the analysis of more than 200 dictionary word entries containing the prefix *wa-*, and on the other hand, on the exhaustive review of 14 texts. Those texts contain 177 forms with *wa-*, including conjugated verbs. However, in almost half of the dictionary entries, *wa-* appears on bound roots or old formations where it has no function in synchrony, like *wakʰéga* 'to be sick'.[2] Such forms were left out of the study.

The different functions recognized synchronically in Umoⁿhoⁿ are:

- 3rd person plural object marker (gloss: O.3PL)
- antipassive marker
  - with a generic object or no object
  - with a referential object
- nominalizer

Whatever its function, the sequence *wa-* undergoes some frequent morphophological changes in combination with other derivational prefixes, as will be seen in the subsequent examples:

/*wa-í/ → wí
/*wa-á/ → wá
/*wa-u/ → ú

[1] See discussions on the Siouan list in 2002, originated from the paper by Helmbrecht (2002) in IJAL: http://listserv.linguistlist.org/pipermail/siouan/2002-July/subject.html#2158

[2] The bound root of this verb is **kʰega*, not attested alone. The conjugated forms of *wakʰéga* show unambiguously that *wa-* is a prefix apart from this root. It is the same with all the verbs beginning with *wa-* surveyed.

2. Main/basic functions of *wa-* in Umo^{n}ho^{n}

2.1. Third person plural object marker

In examples (5) and (6), the prefix *wa-* serves as an indexation marker for the 3rd person plural object of the verb. In each case, the plural object is definite. Note that *wa-* appears independently of the expression of the object as a noun phrase: in (5) the object "the Dakotas" is expressed in the clause, and in (6) the object is not expressed.

(5) *ó^{n}ba théthu sháo^{n} amá **wa**-tó^{n}be ha, (...).*
day here Dakotas the.PL O.3PL-A1.SG.see DECL.M
'On this day, I have seen the Dakotas (...)' (Dorsey 1890:707.1 / Unázhi^{n}-ska)

(6) *íthae-b-azhi-xti-o^{n} shtewó^{n} **wa**-'ú-bi-ama.*
talk-PX-NEG-INTENS-AUX even so O.3PL-wound-PX-NARR
'Without speaking at all, he wounded them.' (Dorsey 1890:361.9 / Joseph LaFlesche)

In Umó^{n}ho^{n}, *wa-* can also express the 1st person plural object (O.1PL), as in the form *wadó^{n}ba* 'she sees us' (Saunsoci & Eschenberg 2016 / Alice Saunsoci). However, O.1PL can also be marked by an additional *a-* along with *wa-*. This is the case in example (7), where O.1PL is marked by *awa-*. According to Carter et al. (2006:6), the O.1PL marker takes the form *a-wá-* each time a preverbal form occurs before it, that is, whenever *wa-* and the other indexation markers are not verb initial.

(7) *égithe uthéwi^{n}-awá-tha-i.*
finally assembled-O.1PL-CAUS-PX
'At length they assembled us' (Dorsey 1890:435.3 / Pathi^{n}-No^{n}pázhi)

2.2. Unspecified argument marker with generic reading: Antipassive

Example (8) shows the common transitive verb *that^{h}é* 'to eat it', with *hébe* 'piece' as an object.

(8) *hébe that^{h}á ga!*
piece eat IMP.M
'Eat a piece!' (Saunsoci & Eschenberg 2016:94 / Alice Saunsoci)

In Umo^{n}ho^{n}, as in other Siouan languages, if no object is expressed on a transitive verb, it is understood as definite: *that^{h}é* means 'to eat it'. The prefix *wa-* saturates the object position and enables the verb to be used without referring to any specific object, like in example (9):

(9) *wa-bthát^{h}e í^{n}-udo^{n}.*
ANTIP-A1.SG.eat DAT.1SG-good
'Eat a piece!' (Saunsoci & Eschenberg 2016:94 / Alice Saunsoci)

In such cases, it functions like an antipassive marker. Antipassive constructions can be defined as intransitive constructions derived from transitive ones with some overt morphological

encoding, where the agent is preserved and the patient either inexpressible or demoted to an oblique function (adapted from Heaton 2017:64).[3]

Table 1: Some antipassive verbs in Umoⁿhoⁿ

Base verb	**Antipassive verb**
baxú 'to write *x*'	*wabaxú* 'to write', 'to write something'
'é 'to farm *x*', 'to hoe *x*	*wa'é* 'to farm'[4]
gí'i 'to give *x* to *y* to help him/her out'	*wé'i*[5] 1. 'to give *x* to folks to help them out' 2. 'to give stuff to *y* to help him/her out'
dóⁿbe 'to see *x*', 'to look at *x*'	*wadóⁿbe* 'to scout'
uhóⁿ 'to cook *x*	*úhoⁿ* 'to cook' (/wa-u/ → [ú-])
thatʰé 'to eat *x*'	*wathátʰe* 'to eat', 'to eat something'

Table 1 shows some antipassive verbs in Umóⁿhoⁿ and their corresponding transitive bases (all forms are common, except when specified in footnotes). Note that the antipassive verb can take a specific, culturally relevant meaning, like in *wadóⁿbe* 'to scout' (widely attested in Dorsey 1890). *Wa-* can also remove any of the objects of a ditransitive verb, as can be seen for *wé'i* and its two possible meanings (UNPS 2015:8).

2.3. Nominalizer

Wa- is attested on deverbal nouns corresponding to intransitive stative verbs, transitive verbs and ditransitive verbs, as can be seen in Table 2. In most cases, the noun corresponds to the patientive argument of the verb, although in some cases, it can correspond to the agent (see last two lines).

Table 2: Deverbal nouns formed with the prefix *wa-*

	Verb	**Noun**
Intransitive Stative	*shíⁿ* '(to be) fat'	*washíⁿ* 'fat', 'bacon'
	zhíⁿga '(to be) small'	*wazhíⁿga* 'bird'
	zhíde '(to be) red'	*wazhíde* 'tomato'
	baxté 'to tie it'	*wabáxte* 'bundle'
	thatʰé 'to eat it	*wathátʰe* 'foot'
Transitive	*áthaha* 'to put it on'	*wáthaha* 'clothes'
	tóⁿ 'to have (many of) it'	*watóⁿ* 'goods'
	baxú 'to write it'	*wabáxu* 'letter', 'writer'[6]
	dóⁿbe 'to see it'	*wadóⁿbe* 'scout'

[3] The antipassive was first described by Silverstein (1972:395) in the description of an ergative language. Since then, many scholars restrict this notion to the description of ergative languages, while others use it for describing object-demoting constructions without considering the alignment features. Here, I follow the second tradition. The Umóⁿhoⁿ antipassive construction described here corresonds to the "functionally determined" antipassive of Cooreman (1994).

[5] *Wa'é* 'to farm' is attested a few times in Dorsey (1890).

[5] *Gí'i* and its antipassive counterpart *wé'i* are attested only in UNPS (2015:8).

Other nouns combine *wa-* with an applicative prefix: the instrumental *í-* 'with', the locative (superessive) *á-* 'on', and the locative (inessive) *u-* 'in' (see morphophonological changes at the end of section 1). Here again, they can be considered as deverbal nouns because the corresponding verb without *wa-* can easily be retrieved. Such verbs are not attested in the corpora available to me, however.

Table 3: Deverbal nouns of instruments and location

Putative Verb	**Noun**
**áthetʰe* 'to eat *x* on *y*'	*wáthatʰe* 'table'
**íbaxu* 'to write *x* with *y*'	*wébaxu* 'pencil'
**ímagixe* 'to carve *x* with *y*'	*wémagixe* 'saw'
**ú'e* 'to farm in *x*'	*ú'e* 'field'

It is not always clear if *wa-* should be considered as a nominalizer, especially when the corresponding verb is transitive: in those cases, it could be considered as a conversion from a transitive verb with antipassive *wa-*.[7] For lack of space I will not cover this issue. I consider that *wa-* has a nominalizing function at least when it derives nouns from intransitive stative verbs. Intransitive stative verbs cannot take an antipassive marker, nor do they take the 3rd person plural object marker when their subject is plural (Marsault 2016:81).

2.4. Temporary conclusion: O.3PL ↔ ANTIP

From the examples in the previous sections, and leaving aside the nominalizing function, we see possible ambiguities between wa- as a 3rd person plural object marker and as an antipassive marker: the antipassive marker could be interpreted as a third person plural maker having acquired a generic reference. Thus, the form *wamóⁿthoⁿ* in (10b) can be interpreted in two different ways, and the context does not always clarify which function *wa-* is taking.

(10) a. *moⁿthóⁿ* 'he steals *x*'
b. ***wa****móⁿthóⁿ* 'he steals (things)', 'thief', 'he steals them'

Example (11) typifies this ambiguity. Here, the form *wéthai*, from the base verb *íthe* 'to find *x*', means 'discovered the presence of enemies', as can be seen in the translation. This example is from the second sentence of a tale, and 'the enemies' have not been previously introduced. Nonetheless, Dorsey glossed the verb 'discovered them', favoring an interpretation of *wa-* as a marker of O.3PL.

(11) *Táxti-zhiⁿga akʰá* ***wé****tha-i* *tʰe.* (< *íthe*: 'to find *x*')
Fawn the discovered them (Dorsey's gloss)
deer-small the.SG ?.find-PX PAST
'The Fawn discovered the presence of enemies' (Dorsey 1890:358.1 / Joseph LaFlesche)

[6] *Wabáxu* is very frequently used for 'letter'. I have found one instance where it refers to the writer, in Dorsey (1890:509.3).

[7] As will be seen in section 3, *wa-* as an underspecified argument marker can also correspond to a referring, singular object which is left indefinite by the speaker. This use of *wa-* could account for nouns such as *wabáxu* 'letter'.

This analysis according to which the antipassive could arise from O.3PL with a generic interpretation works in most cases in Umo^{n}ho^{n}. However, a few corpus examples contradict this analysis, as will be seen in section 3. Moreover, while most Siouan languages have a cognate of *wa-* as an absolutive, only a few of them have the prefix *wa-* as a O.3PL marker.

3. Antipassive with referring object

In example (12), we see the prefix *wa-* on the transitive verb 'to write *x*', where it stands for an inanimate singular referring object. It cannot be interpreted as a O.3PL marker. The gloss provided by Dorsey, "you wrote something", identifies *wa-* as an instance of the function of "underspecified argument marker". In this case, however, it cannot be interpreted as an antipassive marker either. The sequence *washpáxu t^{h}íthathe tho^{n}* forms a relative clause translated 'what you wrote and sent hither' by Dorsey. *Wa-* stands for a specific object which has the syntactic status of a verb argument: whatever *wa-* refers to is the object of the second verb, *t^{h}íthe* 'to send *x*', and it is the head of the relative clause, not expressed as a NP here. Note that the verb is conjugated with A2 personal marker (initial *shp-* instead of *b-*), which unambiguously distinguishes it from a noun.[8] For such cases, *wa-* is considered as an "indefinite object" (glossed INDEF), which does not reduce the verb valency.

(12) ***wa**-shpáxu* *t^{h}í-tha-the* *tho^{n}* *a-nó^{n}'o^{n}* *(...).*
you wrote something you sent it here the ob. I heard it (...) (Dorsey's glss)
ANTIP-A2.write arrive-A2-CAUS the A1.SG-hear
'I have heard what you wrote and sent hither (...)' (Dorsey 1891:64.1 / Gahige)

Example (12) is contradictory with the temporary conclusion presented in §2.4: here no ambiguity is possible with *wa-* as a O.3PL marker, and it suggests that the "underspecified argument marker" is disconnected from the O.3PL marker, despite the fact that they are homonyms.

Furthermore, *wa-* is sometimes attested in clauses where an inanimate or singular object is expressed as an NP, like in (13). To the best of my knowledge, such constructions in Umó^{n}ho^{n} have never been explicitly described, and the translations that come with such constructions do not suggest any interpretation of *wa-*. It is possible that *wa-* has developed a partitive semantics or continuative aspect value. This could account for the presence of *wa-* in (13). With this interpretation, we could propose as literal translations of example (13) as 'He went home after telling us to work some of the land with our hands', or 'He went home after telling us to do some work on the land with our hands'.[9]

(13) *No^{n}bé tó^{n}* *mo^{n}zhó^{n}* *wa-thíto^{n}* *wagázhi* *agtha-í.*
hand have land ANTIP-work O.1PL-ask go.home-PX
'He went home after telling us to work the land with our hands' (Dorsey 1890:507.7 / Teúko^{n}ha)

[8] If the verb were not conjugated, the sequence *wabáxu t^{h}í-tha-the tho^{n}* could be interpreted as 'the letter which you sent here', due to the fact that *wabáxu* can be a verb and a noun (compare Table 1 and Table 2).

[9] This interpretation of *wa-* was suggested to me by B. Gordon (p.c.), with the verb *wathát^{h}e*. They reckons that it needs to be checked with speakers, and I try here to apply this interpretation to example (13).

Considering the previous examples, the precise reference of *wa-* in examples such as (14) is ambiguous: it could be analyzed either as a generic *wa-*, translated as 'I write to you', or as a referring *wa-*, which is the interpretation favored by Dorsey here: 'I write something to you', with the speaker having something specific in mind that will be made explicit in the following sentences (example (14) corresponds to the first sentence of a letter).

(14) *K^{h}agé-ha,* ***wa****-wí-paxu*
friend-VOC.M ANTIP-A1.SG/P2-A1.SG.write
'O friend, I write to you about something' (Dorsey 1890:55.1 / No^{n}zadazhi)

4. Conclusion

This paper identifies at least 4 functions for the prefix *wa-* , with possible ambiguities between them: third person animate plural object marker (O.3PL, §2.1) ; antipassive marker (generic reading, ANTIP, §2.2); indefinite object marker (specific reading, INDEF, §3); and nominalizer (NMLZ, §2.3). Few descriptions of Siouan languages address the issue of the referentiality of the patientive participant in constructions with *wa-*, and it cannot be easily deduced from the English translation. It seems that the generic referentiality is attested in most Siouan languages, at least on a few verbs each time (see Marsault forthcoming).

In Umó^{n}ho^{n}, while *wa-* as a generic antipassive can be linked to O.3PL marker, suggesting at first sight some historical link between them, there is no link between the O.3PL marker and the INDEF marker, which often stands for non plural and/or inanimate objects.

Ambiguities are possible between O.3PL and generic antipassive (examples 10 and 11), and between generic and referring antipassive (example 14). In turn, the functions of referring antipassive (*verb* something) and nominalizer (something that is *verb*) are close, especially considering the pervasiveness of relative clauses in Umo^{n}ho^{n}, and its structure (see Rudin 1991).

Figure 1: Ordered functions for *wa-* in Umo^{n}ho^{n}

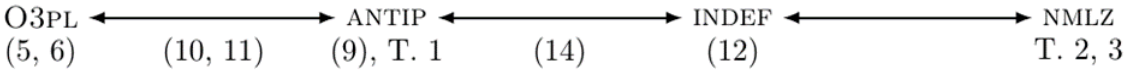


Figure 1 captures the distinct functions of *wa-* described in this paper, and the possible ambiguities between them. The reference of the illustrating example(s) or Table(s) of each function are indicated underneath. The arrows between each function illustrate possible ambiguities between two functions, and the examples illustrating these ambiguities are referred to underneath the arrows. This figure explains why it is difficult to establish a diachronic source for the antipassive in Umó^{n}ho^{n}: the data show ambiguities in two opposite ways. In Marsault (forthcoming), I suggest that the antipassive reading can come from both O.3PL and INDEF, building evidence on formal distinctions which are found in some contexts.

More research is needed on this subject in order, on the one hand, to better understand which are the values of *wa-* synchronically (especially the possible "partitive" value), and on the other hand, to investigate the possible diachronic pathways of evolution, and confirm or infirm the hypothesis of two sources for the antipassive.

Abbreviations

1, 2, 3	1st, 2nd and 3rd person	INTENS	intensive
A	agentive	NARR	narrative
ANTIP	antipassive	NEG	negation
AUX	auxiliary	NMLZ	nominalizer
CAUS	causative	O	object
DAT	dative	PAST	past
DECL	declarative	PL	plural
EVID	evidential	PX	proximate
IMP	imperative	SG	singular
INDEF	indefinite	VOC	vocative

References

Boyle, John P. 2009a. The functions of [*wa-] in Siouan. Accompanying table of the paper "The [*wa-] prefix in the Siouan languages".

Boyle, John P. 2009b. The [*wa-] prefix in the Siouan langauges. Paper presented at the Comparative Siouan Syntax Workshop in Lincoln, NE. June 12-14.

Carter, Richard T., A. Wesley Jones, Robert L. Rankin, John E. Koontz & David S. Rood. 2006. Comparative Siouan dictionary. Unpublished database.

Cooreman, Ann. 1994. A functional typology of antipassives. In Barbara Fox & Paul J. Hopper (eds.), *Voice: Form and function*, 49–88. Philadelphia, PA: John Benjamin.

Dorsey, James Owen. 1890. The Ȼegiha language. In *Contributions to North American ethnology*, vol. 6, 1–794. Washington, DC: U.S. Government.

Dorsey, James Owen. 1891. Omaha and Ponka letters. *Bureau of American Ethnology Bulletin* 11. 1–127.

Graczyk, Randolf. 1991. *Incorporation and cliticization in Crow morphosyntax*. Chicago, IL: University of Chicago dissertation.

Graczyk, Randolph. 2007. *A grammar of Crow*. Lincoln, NE: University of Nebraska Press.

Hartmann, Iren. 2015. Valency classes in Hoocąk (Ho-Chunk). In Andrej Malchukov & Bernard Comrie (eds.), *Valency classes in the world's languages*, vol. 2, 1245–1291. Boston, MA: De Gruyter Mouton.

Heaton, Rania. 2017. *A typology of antipassives, with special reference to Mayan*. Honolulu, HI: University of Hawai'i at Mānoa dissertation.

Helmbrecht, Johannes. 2002. Noun-verb in Winnebago. *International Journal of American Linguistics* 68(1). 1–27.

Kasak, Ryan M. 2019. *Affix ordering and templatic morphology in Mandan*. New Haven, CT: Yale University dissertation.

Marsault, Julie. 2016. *Le système verbal du omaha: Formes et fonctions des préfixes verbaux*. Paris, France: Université de Paris 3 – Sorbonne Nouvelle M.A. thesis.

Marsault, Julie. forthcoming. *Valency-changing operations in umó^{n}ho^{n}: Affixation and incorporation*. Paris, France: Université Sorbonne Nouvelle – Paris 3.

Quintero, Carolyn. 2004. *Osage grammar*. Lincoln, NE.

Rudin, Catherine. 1991. Relative clauses in Omaha-Ponca (Siouan). Paper presented at annual meeting of the American Anthropological Association, Conference on American Indigenous Languages sponsored sessions.

Saunsoci, Alice Fern & Ardis Rachel Eschenberg. 2016. *500+ verbs in Umó^{n}ho^{n} (Omaha): Doing things in the Umó^{n}ho^{n} way.* Scotts Valley, CA: CreateSpace Independent Publishing.

Silverstein, Michael. 1972. Chinook Jargon: Language contact and the problem of multilevel generative systems, part I. *Language* 48(2). 378–406.

Ullrich, Jan. 2008. *New Lakota dictionary*. Bloomington, IN: Lakota Language Consortium.

UNPS. 2015. *Umó^{n}ho^{n} Íye conjugation charts (transivite verbs only)*. Macy, NE: Umó^{n}ho^{n} Language and Culture Center. Bryan J. Gordon (compiler).

More Jiwere-Baxoje fantastic creatures from the Dark Side

Jill D. Greer
Missouri Southern State University

Abstract: This paper continues the inventory of fantastic creatures who were not especially friendly to humans, portrayed in the tales of the Otoe-Missouria and Ioway peoples, as documented in the collections by James Owen Dorsey, Gordon Marsh, Alanson Skinner, and others. Some of these creatures were included in Greer (2019). The supernatural beings include the Great Serpent, geological formations gifted with the ability to think and move, as well as the Dark One. I will explore their attributes specific to these traditions, make comparisons to other Siouan tribes and their folklore for parallels, and utilize the Thompson motif index for North America for further notes on these elements' distribution across Native culture regions. The full set of Fantastic Heroes remains open for future work.

Keywords: Otoe-Missouria, Ioway, Siouan folklore, Thompson motif index

1. Introduction to a generalized Siouan cosmology

While the focus of the paper will be upon stories and lexicon of the Otoe-Missouria and Ioway tribes, I would like to begin with including beliefs also documented among the closely related Hoocąk nation, and the Dhegiha-speaking tribes as well (namely the Osage, Kanza, Quapaw, Omaha, and Ponca). Key elements are shared, not just linguistically, but in terms of general clan structures, subsistence patterns, kinship systems, and folklore. I am also beginning with the latter two groups because they have been linked by many contemporary archeologists with the site of Cahokia, and its area of influence, up into Wisconsin, Minnesota, the Dakotas, Nebraska, Iowa, Kansas, and Missouri in particular, which tie in with the ancestral homelands for most of these peoples.

In the general three-level cosmos documented for the Dhegiha speakers (Fletcher & La Flesche 1911 and Diaz-Granados et al. 2015) and the Hoocąk (Radin 1948), there are three interconnected domains. Above is the Sky World (Sun, Birds, Thunderbirds...), then the Middle World (where humans live, the surface of the Earth), and finally, the Underworld (Darkness, underneath the earth, the place of water, water spirits, snakes, and horned underwater panthers. It is associated with females and fertility.) Upon the surface of the Middle World, there are places that serve as portals to the world below; such liminal places[1] include bodies of water, springs, rivers, and caves.[2]

[1]Liminal is the term used by Victor Turner in his study of human rituals. Part of the action of creating sacred space is to be separated from everyday life, the profane and ordinary. It comes from the Latin root limin or limen, meaning 'threshold' (LEXICO: Oxford English and Spanish Dictionary, Thesaurus, and Spanish to English Translator, https://www.lexico.com/definition/liminal). Thresholds are the physical limit delineating inside vs. outside of a building, being neither fully interior, nor fully exterior, but 'betwixt and between' in location.

[2]Lewis-Williams & Dowson (1988) did seminal work on the importance of shamanistic art on stone, and described

In the Osage interpretation of the Cosmos, there is an Eternal Battle between the forces of Underworld and Sky. I am interpreting LaFlesche and the other authors' complex theology here in a somewhat simplified form, but it seems to me that one implication of that conflict is that humans can be helped or hindered by all powers, and that this conflict is necessary for life on the Middle World to continue—an eternal yin/yang which cannot be resolved. This tension favors the Sky World in many ways but does not split into a true duality of good vs. evil as is found in the Near Eastern traditions such as Zoroastrianism, Judaism, Christianity, and Islam. The Under World brings not only darkness and death, but the Moon, and the feminine powers of fertility and water. Likewise, the Sky World brings the sun and rain, but also the storms, lightning, and fire. Each is necessary, full of power, and terrible in its own right.

Now, we will look specifically at accounts given of Iowa cosmology proper. Long ago, however, things were different. In a mythic time before the historic period/the present, strange and awe-full creatures populated the earth in large numbers and had to be destroyed or reduced in number before puny humans could survive. Foster (1994) summarized traditional Ioway cosmology in his thesis about sacred Bundles. I should have consulted him before doing my paper last year, but I am happy to note that his conclusions based on the earliest accounts are quite similar. Let me share a quote here:

> "The destructive forces were the Underworld Powers as represented by the *ischéxi*, the horned water panther or serpent, ghosts, monsters, giants, and little people (though these last could be good). On top of all these were many unnamed *wakándas* dwelling in bluffs, water, timber, high rocks, mounds, and even household utensils (Dorsey 1894; Skinner 1925, 1926). The world was conceived of as being a lodge, as well as being multi-tiered." (Foster 1994).[3]

1.1. Beneath the surface of Earth/Underworld

As humans rely on a certain amount of light for their keen sense of vision, it is not surprising that darkness itself can be feared. We can document a similar attitude for the Baxoje-Jiwere peoples. One example of the attitude toward the dangers of the Under World is found in the writings of Whitman and Skinner. Both Ioway and Otoe-Missouria elders repeated warnings from the old people for the young people who went on their Puberty Fast alone in the wilderness: "[D]o not speak to creatures coming up from the Water or from under the Ground. They could kill you, you must not speak to them" (Skinner 1915:740)! Consider the following teaching for those about to embark on this sacred venture:

> "Now it is time for you to use the burnt stick (i.e., rub charcoal on your face) and let your tears drop on our mother, the earth, that she may pity you and help you in the future. Find out your way; the creator will help you. He may send a voice to speak to

rock faces as 'membranes' between the spirit world and this plane of human reality; many cultures around the world associate stone with the cold permanence of death, but also linked to the spirit of ancestors. Stanley (2004) documented the spiritual significance of art upon certain rock formations in Iowa, possibly linked to the ancestors of Iowa, Otoe, and Missouria today.

[3] Pertinent information from chapter 3 of Foster's (1994) thesis can also be found online at the following address: http://ioway.nativeweb.org/iowaylibrary/sbchapter3.html

> you and prophesy whether or not you will be of any account in the tribe. Maybe you'll dream of the thunder or some other one above, one of its assistants or servants. They may give you long life. Weep for help from the sun. The sun is a great power.
>
> ***If something comes up out of the water or the earth, don't accept it. Throw it away. Pay no attention to it. Don't listen at all or you'll soon die. That is the way to do. Be careful, there are both heavenly and evil powers, and the latter will try to deceive you.*** You must be willing to fast, for, if Wakanda helps you, you will be a great man and a protector of your people. You will become famous." (Skinner 1925:739-740) (Emphasis added).

1.1.1. Serpents and taboos

Snakes move easily underground and upon water as well as land. This speed and ease of transition between domains, plus the venomous powers of certain species, makes them very *wakan* 'mysterious, sacred, powerful' indeed! Note there were normally strong taboos around killing any kind of serpent (Marsh n.d., 'The Twins').[4] In one historic account, the Anglo-American missionary makes note of an occasion when some children encountered a snake on a trail away from the village. They hurried home, so that the priest/shaman could come back, and offer the serpent some tobacco. This sacred offering would help ensure the safety of tribal members as they travelled and possibly encountered this sacred creature in the future (Hamilton n.d.).

Secondly, there is a direct connection between the proper time to tell sacred myths known as *wekan* and the season of snake hibernation. Telling *wekan* in the 'summer season' is taboo because it would attract these dangerous creatures to one's abode, or the next day, one might step on a snake (Meeker 1901:164). The Hidatsa had another association between snakes, the earth, and women, which was that only those families who had the Snake bundle 1) knew how to make pottery and 2) were allowed or sanctioned to do so (Bowers 1992:373 and Duncan & Diaz-Granados 2015:102).

The mysterious appearance and disappearance of these legless reptiles shows their ability to literally navigate between the Middle World and the Under World. Finally, the rather unique trait of snakes shedding their skins represents new life and eternal rebirth or resurrection. But these snakes are ordinary creatures, not limited to mythic times or supernatural appearances.

1.1.2. The Great Serpent

In addition to the ordinary snakes, there are also tales of a mythical World Snake or Great Serpent, who is a consort with the Old-Woman-Who-Never-Dies (also known as First Woman).[5] She slept with this powerful creature first, then afterward, she also slept with the Sun (also known as First Man). Thus, the sacred woman's bodily cave acts as the literal conduit through which she will transmit the underworld power to the sky world also. This belief in a holy woman's vagina as a

[4]This is direct contrast to the Hero Twins' lack of fear /respect; they cook them up, make rattling curtains for a doorway, etc. (Marsh n.d. 'The Twins' and Skinner 1925:429).

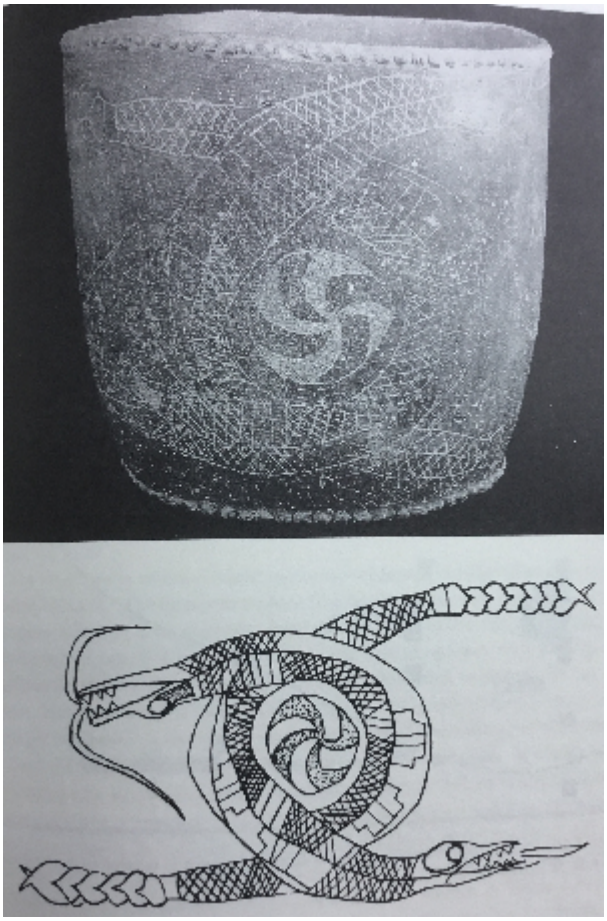
[5]The feminine balance to the masculine warrior/sun/Upper World power is widespread in Native North America. Note that Mississippian figures such as the Birger sculpture depict a female deity who is hoeing the back of the Great Serpent, and has squash growing up her back. Her lips are parted, in the stylized depiction of death, signifying that she is also the keeper of the souls of the dead. These fascinating images are crafted of catlinite, the same sacred red stone used for pipes (Prentice 1986).

medium for acquiring and passing on power parallels a custom found among the Mandan and Hidatsa. A wife might ceremonially mate with an older Warrior/Doctor at the request of her husband; then, after that ritual sexual union took place, her husband would then acquire that male elder's sacred power when the husband was reunited sexually with his wife again (Bowers 1992 and Peters 2000:40-41).

The Hoocąk Medicine Dance was a tribal religious society found among the Otoe-Missouria, Ioway, and many non-Siouan Midwestern neighboring tribes as well. Radin (1950) noted that members of the Hoocąk Medicine Dance used poetic names for snakes: "Spirit Walking Soldiers" and "Crawling Soldiers". The reasoning was to avoid their hearing/noticing the use of their true name during a ceremony, and thereby being naturally curious enough to be summoned to the event. Similar poetic names might be used ceremonially for other powerful creatures, especially the bear.

The Osage had a Rite of Reincarnation for the Dead that invokes the Great Snake. This sacred being, the Great Snake is quoted in the ritual as saying "[e]ven though the little ones pass into the realms of spirits, they shall, by clinging to me and using my strength, recover consciousness" (La Flesche 1975:368, quoted in Diaz-Granados et al. 2001:488).

Figure 1: Engraved rattlesnakes intertwined around cosmos on pot, Cagle Lake, Missouri (O'Brien 1994:180-181)[6]



[6]I am using this image as a generic example of the frequent imagery of ordinary snakes, especially venomous

The Kanza had a mythical flying snake called *wéts'a tazhi lishka*. It was so powerful that just seeing it would mean that person must die. Note also that there was an old woman who used to hold supernatural communication with them (Cumberland & Rankin 2012:214; James Owen Dorsey is their original source). Compare this mythic being to the imagery of a Winged Serpent engraved upon Mississippian ceramic vessels such as those found at the site of Moundville, Alabama (Reilly 2015:140). Look at the following illustrations. The first is an engraved jar in Fig. 1 above. Figure 2 is from the incredible Spiro Mounds site near present day Ft. Smith, Arkansas. The images are included, not because we want to directly link the Kanza with these specific sites, but to document the widespread nature of this belief system, and its roots in the overall Mississippian multicultural and poly-linguistic complex (cf. Kaufman (2014) for a study of linguistic diversity in the southern Mississippian region).

Figure 2: Hemphill-style Winged Serpent Pot, Moundsville, Alabama (Beam 2018)

Now that these various associations of snakes and the underworld have been established, what about the liminal places such as caves /chasms, etc.? They have a certain amount of ambiguity,

ones such as rattlesnakes and copperheads. However, the stepped design in the center of their backs, and the circular motif in the center may actually denote the sacred four snakes put by Earth-maker as anchors of the earth at its four corners, to stop it from spinning (Radin 1950). Since the stepped design and the circular motif (there defined as the cosmos) are both found in the Osage symbols still recognized in the 20th century (Burns 1994), I am speculating on the interpretation. O'Brien (1994) does not interpret this artifact in symbolic terms.

since the Earth is our Grandmother, but they also are portals to the third dark and terrible realm, which is also powerful, and brings a necessary balance for the cosmos, like day and night, rain and the roots of plants under the soil.

Cracks in the Earth and caves being perceived as portals to that Underworld are extremely widespread concepts. How might this relate to the practice of ritual, and rock art? Perhaps best known is the fascinating work by David Lewis-Williams finding parallels between San (South African) shamanic rock paintings as well as Upper Paleolithic cave art in Europe. Remember that complete sensory deprivation such as one finds in total darkness and silence deep inside cave passages is one avenue for entering into a trance state, suitable for spiritual seeking.[7]

The topic of chasms and caves brings us to our next unusual fantastic being, which is neither anthropomorphic nor of the avian/animal world. Rather, it is a part of the geology of the world itself. It is quite similar to the way humans have viewed volcanoes as sacred and living creatures, too.

2. Geologic monsters

2.1. Hill That Swallows Living Creatures (humans, sometimes animals, too)[8]

2.1.1. Swallowing Earth/the Devourer

Considering caves, chasms, perhaps even landslides and earthquakes, we may be able to relate this important motif to actual specific geological features or processes, including the very real dangers of the many caves all over North America. In his survey of Native American narratives, Thompson categorized this recurring story element as "*#158. Sucking Monster (G332). Giant, sometimes represented as a giant hall or cave, sucks in victims... includes versions found for the Crow, Hidatsa, Wichita, and Pawnee. To avoid repetition, since most are also on #159, the majority are listed there: #159 Monster killed from within* (K952). Relevant tribes include Osage, Ponca, Dakota, Winnebago, and again the Caddoan-speaking Pawnee (Thompson 1966:321, 364-365).

There is not a single shared name for this monster, which appears in three of our major Jiwere-Baxoje folklore collections, both Dorsey in the 19th century, as well as Skinner and Marsh from about 1910-1936.[9] It is interesting to add here that we may be able to tie this tale to a particular known cave. The Ponca were said to have found the Wind Cave in the Black Hills, and they called it ***Pah-hah-wah-tha-hu-ni***, 'the place that sucks in/the hill that swallows in' (Howard 1995:20).[10] Today the cave is a national park today, in part because of the unique characteristic of air flow

[7]Picture Cave includes 'deep cavern' work also, but there are signs of torches having been regularly used there, whether for the original artist to see, or for initiates to witness the sacred stories illustrated on the rock walls, it is hard to say. But the association between walls of a deep cave and contact with the spiritual world is still relevant, either way (Diaz-Granados et al. 2015).

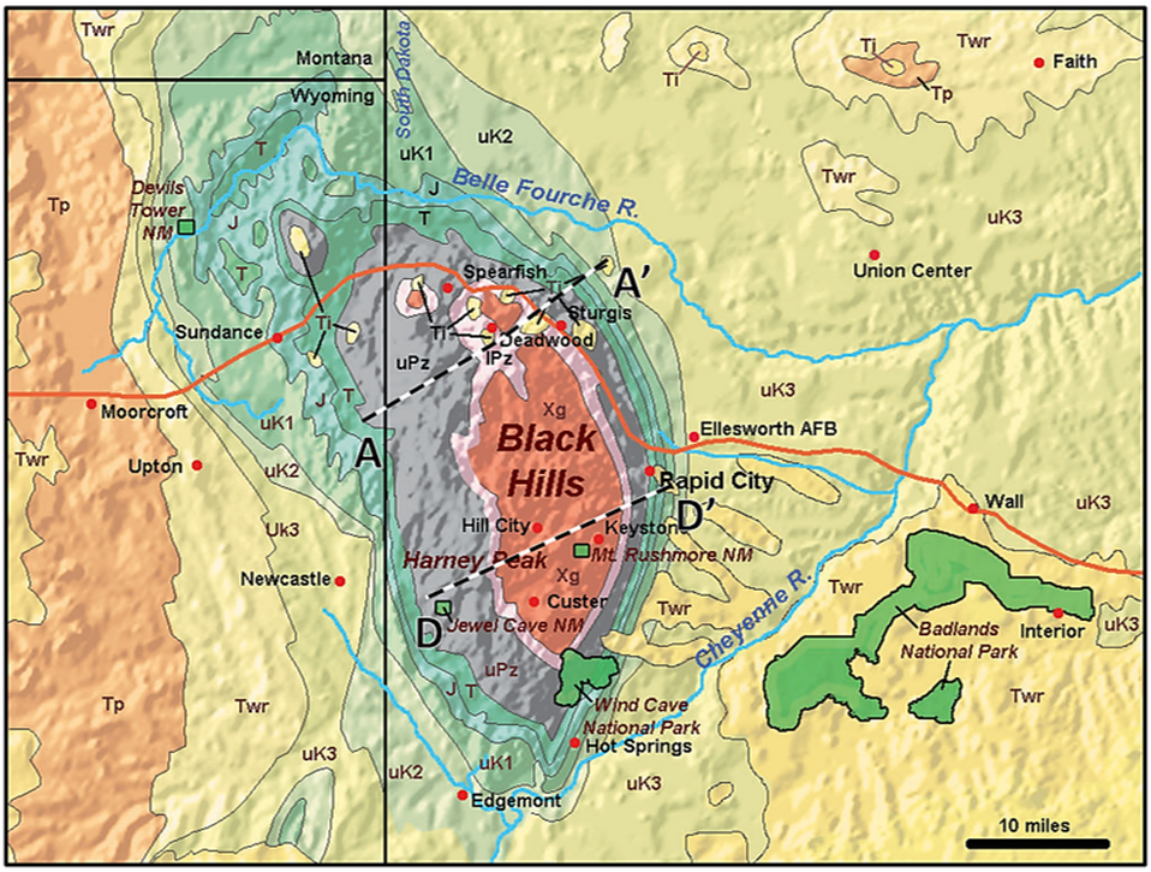
[8]Motif #159 also had a subset (a) #F913 *Victims rescued when swallower is killed*, represented by Osage, Ponca, Dakota (Thompson 1966:321).

[9]Skinner collected tales from various Ioway people, but many were from Robert Small, while Marsh relied heavily upon Mrs. Small.

[10]From a manuscript written down in 1949 by an elder southern Ponca man dictating it to a younger Ponca man. A copy was kept in a safety deposit box for safekeeping before he shared it with Howard! His ethnography was published originally ca. 1965.

between the outside air, moving inward (hence 'sucking in'). See Figure 3, illustrating that this unique cave is well within the traditional hunting and travelling range of the IOM, Dhegiha, and many other Siouan groups.

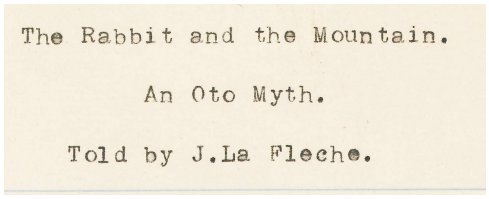
Figure 3: Map of Wind Cave, also known as 'He-sucks-them-in' (www.geowyo.com)



Two versions of this popular motif use the term *Uye* 'Vulva/Vagina', while the others call it 'the Hill that Swallows up (People)' or 'the Devourer'. It is found in both a Trickster story (Hare/Rabbit), as well as the Hero Twins' extended adventures, which are best thought of as a series of connected narratives, similar to the epic journey of Odysseus. At one point, one narrator also calls it "he-sucks-them-in", which parallels the Ponca name for Wind Cave mentioned earlier.

(1) *mąyą kaígi* ***wólahoče****yą* *ítąŋe* *kʰE*
land over.here **he.suck.them.in**-one there-it.is.lying DECL.MALE
'In this land there was a Devourer' [my free translation] ("The Twins", Marsh n.d.:LN 107)

An Otoe Version Takes place in "Rabbit and His Grandmother" (Earth/Old Woman). See example (2) below.

(2) 
The Rabbit and the Mountain.

An Oto Myth.

Told by J.La Fleche.

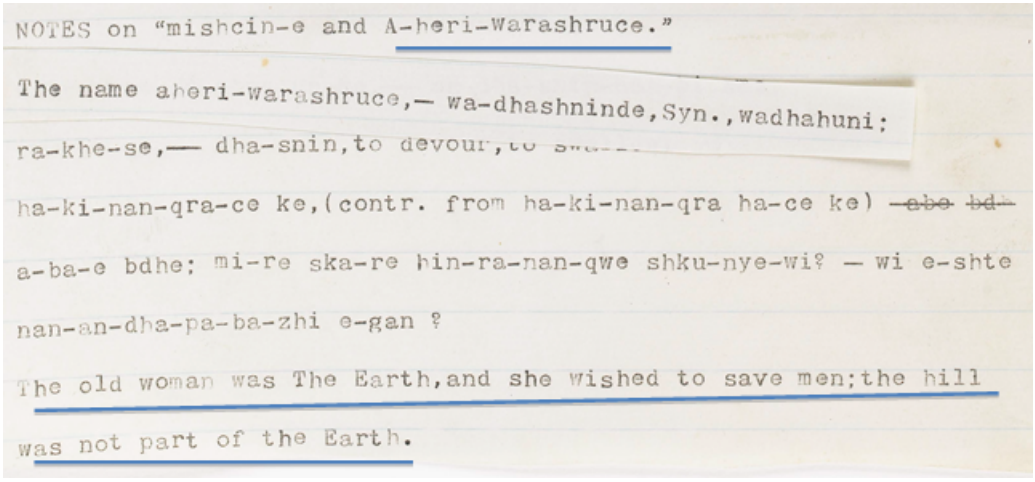
(Dorsey Box 4800 Folder 305)

Aheri	*warashruje*
aheri	wa-ra-shruje
hill/mountain	3PAT-by.mouth-draws.in

'(he) draws them inside by his mouth' (Dorsey n.d.:Folder 304, GoodTracks 2008)

Dorsey gives as synonym ***Aheri wa-dhashninde***, < ***dha-snin*** 'to devour, to swallow', ***wadhahuni***, with no derivation/morpheme by morpheme gloss; he adds in his notes that the monster is NOT part of the earth! (Underlines added by author.)

Figure 4: Scanned image of Dorsey's notes, emphasizing the distinction between Grandmother and the wicked Devourer (Dorsey Folder 304)[11]


NOTES on "mishcin-e and A-heri-Warashruce."

The name aheri-warashruce,— wa-dhashninde,Syn.,wadhahuni;

ra-khe-se,— dha-snin,to devour,to s...

ha-ki-nan-qra-ce ke,(contr. from ha-ki-nan-qra ha-ce ke) ~~-ebe bd-~~

a-ba-e bdhe; mi-re ska-re hin-ra-nan-qwe shku-nye-wi? — wi e-shte

nan-an-dha-pa-ba-zhi e-gan ?

The old woman was The Earth,and she wished to save men;the hill

was not part of the Earth.

Consider the following example also. It is from Dorsey (n.d.:4800, Box 305)[12] again below with an Omaha version given; compare to his notes given earlier on the Hill not being part of Earth![13]

2.1.2. Terms for a geologic monster in related languages

Since J.O. Dorsey was so interested in comparing different Siouan languages as he was studying several of them, we often get direct listing of similar terms in a second or third language, even as

[11] 
naha ǿitceka arenye ki," e. Cike eta Aheri qaⁿtceqtci iyaⁿ itañe.

Dorsey has no glosses here; I cannot tell for certain if this *Aheri qantceqtci* is another name rather than a verb about Rabbit going back to the mountain again, so I will exclude for the time being (*ibid*). However, the fact that iyan follows the lengthy construct suggests it is at least a compound noun here.

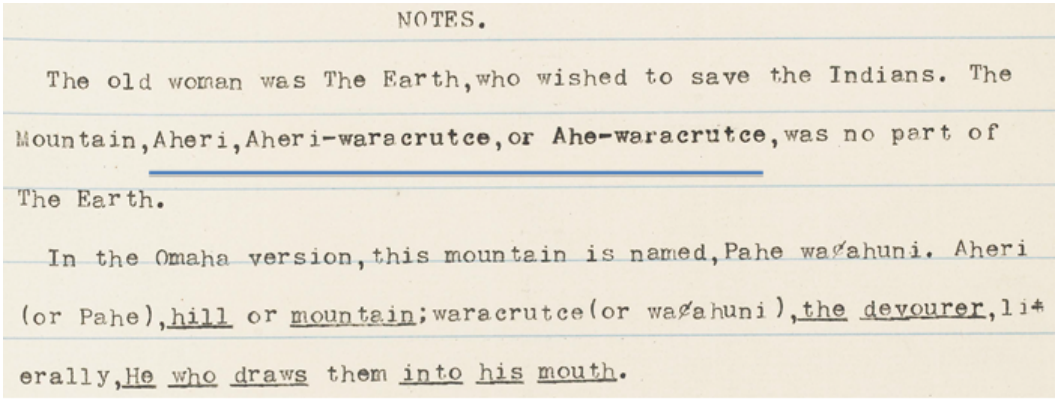
[12] His second drafts preparing to gloss a text are in a different folder, as are the actual translated English versions.

[13] Dorsey is not consistent in his spelling. We have alternates ***wa-racrutce*** and ***wa-rashruce***; possibly an evolution in orthography?

he was trying to focus on translating a story in a different language. Such is the case here, where Dorsey's notes in the Otoe tale go on to discuss what the equivalent words would be in Omaha, too.

See Figure 3 below for the Dorsey's own typed manuscript of his translation for "Rabbit and Grandmother"; the Notes include the Omaha version; compare to notes in the transcription presented earlier about Hill not being part of Grandmother Earth (Dorsey n.d.:Folder 305).

Figure 5: More Dorsey emphasis on *Aheri* as separate from Grandmother Earth (Dorsey n.d.: Box 4800, Folder 305)

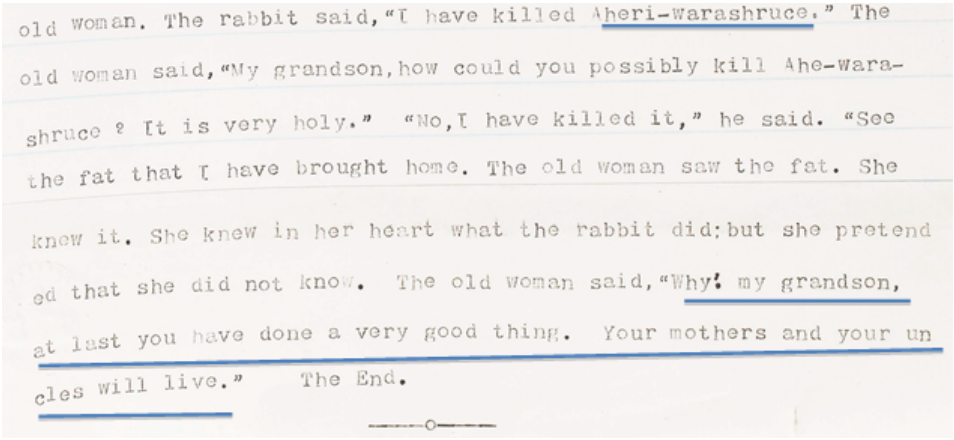


NOTES.

The old woman was The Earth, who wished to save the Indians. The Mountain, Aheri, Aheri-waracrutce, or Ahe-waracrutce, was no part of The Earth.

In the Omaha version, this mountain is named, Pahe waʔahuni. Aheri (or Pahe), hill or mountain; waracrutce (or waʔahuni), the devourer, literally, He who draws them into his mouth.

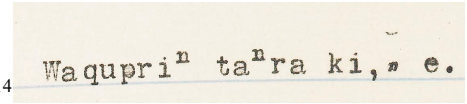
Figure 6: Additional Dorsey emphasis on *Aheri* as separate from Grandmother Earth (Dorsey n.d.: Box 4800, Folder 305)[14]



old woman. The rabbit said, "I have killed Aheri-warashruce." The old woman said, "My grandson, how could you possibly kill Ahe-warashruce? It is very holy." "No, I have killed it," he said. "See the fat that I have brought home. The old woman saw the fat. She knew it. She knew in her heart what the rabbit did; but she pretended that she did not know. The old woman said, "Why! my grandson, at last you have done a very good thing. Your mothers and your uncles will live." The End.

———o———

Folklorists such as Alan Dundes and others like to analyze this one as clear case of Male Birth Envy. Heroes bring forth life from dark living space through their stone tools and heroic actions! Carroll (1992) describes the subconscious feminine danger depicted across many Native American myths about caves that swallow, as related indirectly to the *vagina dentata* motif that

14 
Waqupriⁿ taⁿra ki," e.

"(He) is very holy," she said (Dorsey n.d.: Box 4800, Folder 305).

occurs across both North and South America (Thompson 1966:115). Thus, I love the fact that in the Ioway version, the monster's name is actually *Uye* 'Vagina', because it is not buried in the subconscious at all, but bluntly put right out into the open (Skinner 1925:429-430, 497-498). As the earlier picture of his manuscript clearly shows, Dorsey repeats the consultant's words frequently that the *Uye* is not of this earth, because Grandmother Earth loves all her children and would never wish to harm them (Dorsey n.d.).

In the Ioway version of the same story, Grandmother rescues Hare the first time. The second time, he rolls hot stones down the hill into the cave to kill the *Uye*. This is another motif in Native North American tales, (Thompson 1966:324, Note 167 and 365); Motif #K951. *Monster killed by throwing hot stones into throat*. However, there were no Siouan, Caddoan, or Algonquin tribes mentioned in Thompson's documented examples.

2.1.3. Other Siouan and Caddoan parallels for the Swallowing Monster

In the Hoocąk version actually says it really is Grandmother Earth's *Uze*. She told Rabbit never to play near her genitals, but he disobeyed, and fell inside. It was a cave, full of animals and people. Radin basically translated it as "Womb of the World" (Radin 1948:103-104)! In this version, it is yet another violation of the incest taboo by the trickster. The many ways in which the naughty and foolish Rabbit tricks Grandmother into having sex with him is a frequent theme in these humorous stories illustrating the dire consequences of breaking strict taboos.

We can find additional related motifs: Sucking Monster G332: *Giant (sometimes represented as hall or cave) sucks in victims*; Note #158 includes Crow and Hidatsa, as well as Pawnee and Wichita (Thompson 1929:321). Now, we should add Osage to the list here as well, because in reviewing this piece for our proceedings, Justin McBride generously shared the Osage version of this story with me, called "the Devouring Mountain." It comes from the Osage texts collected by the same prolific fieldworker James Owen Dorsey (n.d.). In the Osage version, humans are sacrificing young women to appease the hungry mountain, and it is the Orphan who bravely decides to stop this dreadful practice by killing the Devouring Mountain himself and winning himself a lovely wife in the end. Note that the Orphan is another heroic figure shared with the Otoe-Missouria and Ioway traditions, although he is not credited with this particular adventure in the existing record of their lore. I suspect that a wider survey of all the Siouan mythic material would yield many additional parallels beyond those mentioned in this preliminary work.

2.2. A related theme: More dangerous geologic feature

The edges of the flat earth are cracks that open and shut. Different heroes ventured across these boundaries at the edge of the earth, one at each of the cardinal directions. It is clear in the stories that humans or animals might easily fall into the openings and have an awful, if unclear, fate. However there was also a Guardian being at each cardinal direction point also, but these guardians are portrayed as basically positive (Skinner 1925:438 and Marsh n.d.)

2.3. Burial mounds: Ghost brings an earthen feature to life

In this instance, we have a human-made geologic feature, rather than a naturally occurring one. However, since mounds have been being constructed in various forms on this continent for several

thousand years, it is possible that their antiquity may have shrouded (!) that human origin but maintained the spiritual nature and supernatural power associated with them. It is about the culture hero Rabbit (or Hare) and another of his many misadventures. He was alone and hungry, wandering over the land, when he discovered a Burial Mound upon which he takes refuge. There is a big taboo being broken here; the living ought never disturb the dead, especially those to whom one is not related. To Hare's surprise, the mound's ghost speaks and miraculously gives food to him. At first, he is grateful, but as he recovers, he notices the putrid odor emanating from the mound. Hare insults it by nicknaming it 'Death-Smell' so it chases him in anger and nearly catches him. It pulls his tail off, in fact (Skinner 1926:501, Tale #45).[15] The imagery of a mobile and agile burial mound is quite striking and unusual. In fact, there is no such motif in the entire Thompson corpus as originally published (1966).

3. More darkness: Devil and dragon

3.1. The Evil Spirit/No-Good God

This particular topic is a difficult one, because it so obviously has parallels with the Judeo-Christian tradition of Satan. Christian missionaries needed to find a way to translate the fallen angel Lucifer into native languages in order to proselytize their congregations. That need opens the possibility that we are dealing with lexical constructions that may be relatively recent coinages, especially if used as a proper noun or name.[16] However, a modern Christian theological usage does not negate the existence of pre-contact religious concepts. For example, *Wakąnda* may be used to convey the Christian concept 'God', yet it stems from an ancient root meaning 'sacred/mysterious/powerful', and it is found in many traditional genres of music, including War Dance songs, as well as in prayers. In addition, as many elders sincerely communicated, they also had older terms such as *Mą-'ų* 'Earth-Maker' that conveyed the notion of a Creator deity, long before Europeans arrived with their Bibles and worldviews. Such beliefs do not presuppose monotheism per se, but did perhaps provide parallels that lent themself to incorporating the new religion with indigenous cosmologies.

Likewise, just because a term was historically used to represent Satan, does not automatically mean that the same term or something similar had not been present before contact, or that there was no conception of negative forces or evil in the world in traditional stories. With that cautionary note, and recognizing that it may be impossible to ever settle such issues satisfactorily,

[15] Earth effigies are found frequently in the regions where Jiwere-Baxoje speakers lived, including Oneota archeological sites, both on a large scale, but also small ones. They might have served as visual markers of clan identity and ownership, according to Radin (1911:520-528). There are Oneota and even Archaic era echoes in this tale, because there are not just burial mounds (conical in shape), but lots of effigies, from Kansas all the way to Ohio! It is also noteworthy that the Kanza had a clan called the Earth Workers Clan *Manyínka Gághe* (Cumberland & Rankin 2012:279).

[16] French Jesuits established the first mission ca.1720 at the trading outpost called Ft. Orleans on the Missouri River near present day Marshall, Missouri (Bray 1961:216-218). Ft. Orleans would have likely influenced both the Missouria and the nearby Osage village. Baptist and Presbyterian missionaries were present from the 1830s on, including Revs. William Hamilton and Samuel Irvin, whose publications on the Ioway language included a grammar, a hymnal of sorts, and a catechism. Their orphanage provided a home but also an assimilation environment for learning English, and Christianity. Our consultant Truman Dailey knew his maternal grandmother well. She had been sent to the Highland Mission school after her father drowned in the Missouri River, but she had fond memories of her time there (Stanley 1990; author's fieldnotes).

since there have been no Ioway and Otoe-Missouria peoples isolated from Euro-American religious beliefs for over 190 years, let us consider the occasional 'dark' or 'evil' figure occurring in otherwise traditional stories.

(3) *Wakʰąda pʰi skųñį*
god/spirit good -not
'No-Good God' (my free translation) *IOM*

Wanathuje is that famous prophet. This culture hero meets 'the Evil One' (in human-like form?) who verbally threatens him near the Mississippi River. The hero is wisely cautious of the dark power, so he goes home by a different route, to avoid contacting the evil one again (Skinner 1926). It might be important to note that *Wanathuje* actually parallels Christ in some ways since the hero tale states he was born of a virgin, and that he chose to come to the Ioways because they had the good sense to keep menstruating women separate from everyone else. Their 'alone-houses' protect the rest of the tribe from the power and danger represented by women's overt sign of fertility, which made them 'cleaner' and holier than many other Indian peoples (according to the Ioway legend). There is a dangerous encounter with the Dark One, roughly paralleling the temptation of Christ in the wilderness. However, *Wanathuje* is also a warrior, husband, and father, who lives a VERY long life, rather than being a non-violent teacher, sentenced to death, placed in a tomb, then miraculously resurrected, so there are distinctly native aspects to his heroic identity as well (Skinner 1925:478-479).

Stories told and passed on primarily through an oral tradition, rather than a written one, do exhibit variation and change over time. Bits of information retained from sermons might filter their way into old stories, and become transformed, or vice versa. One of the frightening abilities the 'No-Good God' has is that he knows what everyone does, and hears what they say, even if it is whispered (Skinner 1925:444). That uncanny characteristic is certainly shared with the Judeo-Christian devil, although sorcerers and other beings may sometimes have had such powers also.

Perhaps related to the concept of an evil god, I found mention of another evil spirit in "The Twins" (Marsh n.d., Line 282) but with a different name:

(4) *wanąxi pʰiškųñį*
spirit (be)good-NEG
'No-Good Spirit/Evil Spirit/Demon' [my free translation] (Marsh n.d.:LN 282)

In the tale, a bad spirit (possibly a ghost) tries to kill Twins by making a sweatlodge too hot, but the wild Twin pours on too much water as the boys hide under a mussel shell. They triumph over that evil spirit, who runs away (Marsh n.d., "The Twins"). However, in Skinner's version, the narrator went on to say that the boys "drove him into the next world, where he remains invisible, but evil. He is the evil one and knows whatever we do or even whisper. He is one of the tribe of Ghosts (*Wanagri*)" (Skinner 1925:440).

3.1.1. Parallels with other Siouan tribes of an Evil One

Among their closest linguistic relatives, the Hoocąk, we see a few parallels, but nothing identical (unlike terms like Earth-Maker and *Wakąnda*). Miner gave the Hoocąk term *waxopįnįšišik* 'devil' in his unpublished lexicon (Miner n.d.). While he did not analyze the form into its basic parts, I will attempt to do so here. The first word *waxopį* appears cognate with the Ioway-Otoe-Missouria (IOM)

word *waxobrį*, *waxobį* meaning 'sacred/holy/dangerous' (applied to shamans/shape-shifters). The *pį* is a widespread cognate meaning 'good', with the oral vowel nasalized by the potential spread of nasalization across all a word's syllables, from right to left, and vice versa (Kasak & Lundquist 2019).[17] The final portion *nįšišik* appears to be the morpheme signaling negation plus the diminutive (*nį* + *šįŋe* 'little' in IOM). The two syllable word 'little' has of course been first reduced to single syllable *šik*, ending in a stop of the same place of articulation in the sister language, Hoocąk. Finally the single syllable was reduplicated, perhaps to indicate that there are numerous less powerful evil spirits about in the world.

A second potential parallel in Hoocąk is their name for 'Devil's Lake', but other than the word *waką* 'sacred/holy/powerful', I will not try to break it down further. *Teewákącąk ('eeja)* 'Devil's Lake/Sacred Lake', would offer tobacco there' (Helmbrecht & Lehmann 2010:192). Devil's Lake is the modern English name given by whites, associated with Native offerings of tobacco in order to cross the water safely. It is true that all the tribes considered it a sacred /mysterious site, but the monsters therein were Native, not Judeo-Christian ones. Note that the Hoocąk consultant quoted in late 20th century narratives didn't have the aforementioned negative view of the underground and underwater powers (Hartmann & Marschke 2010:66-67). This person made no mention of any Faustian "catch" in accepting such power.[18] The three versions of their creation myth documented by Radin (1950) included many evil spirits, but they were distributed among all three Worlds, not just the underworld; in addition, the water-spirits under the earth were considered positive. The necessity of water for life no doubt underlies this aspect of the cosmology.

Folklorist Dorothy Brown collected a Hoocąk story of a great primordial battle between the Thunderbirds of the Sky forces vs. the 'Water spirits, or underwater panthers' (Jarrell & Farmer 2019). The battle is what caused the rugged rock formations surrounding the lake, and not all of the water monsters were killed. Some survived and still live there (*ibid*)!

While the 'Devil's Lake' is yet another example of things being 'lost' in translation between the two cultures, there is one similarity yet again in many traditions, namely the Faust-like bargain these underwater monsters might make. For instance, an Omaha man was fasting alone in the traditional manner, seeking a vision for spiritual power while out in nature, when such a creature appeared to him from beneath the water. It offered him a long life and great spiritual power, which he found appealing. The account goes on to warn that there was always a tragic catch to any gift received from a dark power, which would be hidden from the human recipient until it is too late. Usually it involved the death of a family member(s) (*ibid*). Dahlstrom (2003) also documented this theme for a Fox tale.[19] An Otoe man told how he had been approached by an evil spirit, who offered him lifelong success in hunting deer. But in the end, this attractive power cost him his family, much to his regret (Whitman 1936).

[17]Note that certain consonants may prevent that nasal spread, such as a glottal stop ['], and other stops such as [k] (Kasak & Lundquist 2019).

[18]A number of the Wisconsin Hoocąk are still members of the Medicine Dance or Lodge, which may well have kept their belief system more insulated from Western influences, or conversely, this may represent a modern revision, in order to deliberately reject any potential parallels with Judeo-Christian cosmology. The 'traditionals' tend to keep separate from the members of the Native American Church, and vice versa.

[19]The Mesquaki Scroll Story had parents die; a winged person prevented death of the man with a bow, driving away the horned monster when it returned to take his life, too. The winged one said that the parents' deaths were already payment enough (Dahlstrom 2003)!

Figure 7: Map of Devil's Lake, Wisconsin[20]

A third name may relate to a native concept of a primordial Evil One, from the prolific collected texts transcribed by Radin in his early 20th century fieldwork. That word is capitalized as a proper name, just as he does for Earth-maker, and Hare, Turtle, etc. Unfortunately, he gives no gloss of the name, and I have not seen this name in any other documentation: "...***Herecgunina*** and his attendants, that all the evil spirits and their helpers - those on earth, those from above, and those from below, as may as there were that had sharp teeth, as many as there were that had sharp claws, indeed, as many (evil spirits) as existed - that they were at work upon us" (Radin 1950: Line 80). I will not attempt a morphemic analysis, other than to observe a possible negation lexeme *-cguni-* 'not' (cognate to IOM) In modern orthography it would be *škųnį* 'not; (be) not').

In another example, from stories told to me about the Native American Church, all members are cautioned from going out of the tipi during the all night service and wandering away from the spiritual safety of the group and the fire. One man who did not heed that warning is said to have met an evil being out there in the darkness. The 'devil' tempted him with the vices of cards and gambling, as well as with a promise of the sexual favors of women in the future (Truman Dailey in JDG fieldnotes).

(5) Some Dhegiha groups' terms for the Dark One

a. Kanza: *wakánda pízhi (ni)* 'devil' (Cumberland & Rankin 2012:275)
b. Ponca: *Wakánda péži* 'the bad god' (Howard 1995)

Howard also mentions that Dorsey (1894:371) gives the identical term but Dorsey believes they made this word 'after they learned of him from the whites' (1995:99). In a 1949 interview, one Ponca man explained they had always been monotheistic, but there were different ways to honor God. From the pipe ceremonies, War Dances, and Sun Dance to peyote church or Christian Church, all these ways were good, and it was the same God being honored by all. He went on to say: "*Wakánda-péži* is the same as the Devil to the Poncas. He is the bad god, and seeks to lead men into evil ways. There are other spirits or demons, but there is only one real God..." (Howard 1995:99).

[20]This map is from an article of the Chicago Tribune from https://tinyurl.com/ybq845td.

3.2. There be dragons?

Whitman described an Otoe-Missouria tradition tied to a specific dangerous location, where it was known that a Horned being with seven heads sometimes arises out of the Mississippi River. The elders to whom Whitman spoke included those who had made the long trip from their original reservation in southeastern Nebraska. Thus, I would wager that their recollections were based on firsthand knowledge of the Missouri River, and hearing the stories associated with different places along that long, mighty body of water. Whitman did his fieldwork during the period 1935-36, which means that it had been more than 50 years since the Otoes had lived on the Big Blue Reservation up north (Whitman 1936). The telling and re-telling of such specific geographically linked tales suggests the importance of these beliefs to traditional members.

Thompson likewise discovered another multi-headed creature, which he curiously decides to call by the European term 'dragon': #289 LXXVIII. **The Seven-Headed Dragon**, with Osage, Ponca, Assiniboine, and Biloxi listed as Siouan groups who shared this motif (1966:358). A more general motif #287 *Additional Motifs, (f) Many-headed monsters* (B15.I) has no Siouan or Caddoan tribes listed, but he did include Northeastern tribes having *Horned serpent* under this same heading (Thompson 1966:357).

Lance Foster (p.c.) feels this is a clear Christian borrowing, from the Book of Revelations; he likewise considered the concept of human-like beings with wings to be a Judeo-Christian borrowing, plus Satan/Lucifer as the evil or no-good god. However, there are at least four winged little men with great power, Thunder-man, Lightning-man, Rain-man, and Little-god, described in the Hero cycle too (Skinner 1925:439; Marsh n.d.) Foster may well be correct, but at the very least, a general concept of evil beings 'hell-bent' on harming humankind appears to be of older status, because such beings play a central role in different culture hero cycles.

The horned serpent also had quite specific geographic associations with a great river. Consider the famous lost pre-contact rock art of the sacred and dangerous Piasa, which once overlooked the Illinois bluffs. It was documented by Marquette and Joliet on their voyage down the Mississippi River in 1763 (Reilly 2015:141). See Figure 8 for a line drawing based on their historic sketches. The priests' native guides averted their eyes rather than look upon the powerful monster's large image. Note there are similarities with the fearsome Horned Water Monster, especially the horns, the frontward facing of the head, and the very long tail (cf. Greer 2019 for an overview of the Horned Water Monster in IOM lore). Versions of the Piasa are also documented on artifacts from Spiro Mounds in Oklahoma.

As for 'angels' being necessarily derived from Judeo-Christian influence, I found evidence potentially contradicting that idea. First, the Radin mentioned positive spiritual beings who were winged in his The Origin of the Medicine Rite among the Hoocąk (1950). Secondly, while it is further away in time and space, there is a fascinating artifact that might dispel that presumption that all winged anthropomorphic image must be of European influence. It is a Hopewell-era greenstone effigy pipe of a human-like figure with wings. Hopewell culture is slightly earlier and further to the northeast than the Mississippian complex in the earlier figures, in the A.D. 400-800 period. I suggest this elegant pipe in fact represents mythic beings such as Thunder-man and Little-god. (See Figure 9.) Note that it is commonly considered an image representing a shaman, since there is a widespread belief that shamans' souls can leave their bodies, and fly to the spirit world. Alternate interpretations are always possible.

Figure 8: Reproduction of Piasa Pictograph as described by Marquette & Jolliet[21]

Figure 9: Human Effigy Pipe of Winged Shaman, Ohio[22]

4. Conclusion

Beyond the joy of exploring these rich oral narratives, I hope to have demonstrated the intertribal nature of the tales, and illustrated the deep ties each people had to their specific ancestral homelands, whether it be the Cave that Devours near the Black Hills, or the Devils' Lake in Wisconsin. Furthermore, I hope to continue to look for the interplay of Jiwere language with material culture, from both ethnographic and archeological sources, as begun in Greer (2019), but which is really a continuation of the traditional Boasian unified four-field approach to anthropology. I found inspiration from archeologists who were integrating the study of Siouan (especially Dhegihan) folklore with their interpretations of upper Mississippian artifacts and cave art, beginning with the work of Duncan & Diaz-Granados (2000) and their associates Diaz-Granados et al. (2015). It is just one

[21]This picture is from https://www.facebook.com/pg/Southwestern-CUSD-9-457567940940343/posts/ cf Reilly (2015:141).

[22]Drawing by Richard Balthazar from Pre-Columbian art at www.richardbalthazar.com.

thread in the complex tapestry of interactions between the Siouan groups and their Algonquin neighbors, and especially their linguistic cousins, the Caddoans, and even the Iroquoians (cf. Chafe 1976 for an overview of the similarities in grammatical categories and relations of these three distinct language families which suggest a deeper time frame or 'super-family'). Due to time constraints, I am once again left with at least one quite significant fantastic creature that cannot be discussed here. The haunting and very widespread tale of Grizzly Woman must wait yet a little longer.

References

Beam, Anna. 2018. Recovered Moundville artifacts. AL Nov. 2018. (Available at https://www.al.com/news/2018/11/3-of-the-hundreds-of- stolen-artifacts-from-moundville-recovered.html, accessed July 2020).

Bowers, Alfred W. 1992. *Hidatsa social and ceremonial organization*. Lincoln, NE: University of Nebraska Press. Reprint of Bulletin 194. Washington, DC: Bureau of American Ethnology. (Originally published in 1965).

Bray, Robert T. 1961. The Missouri Indian tribe in archaeology and history. *Missouri Historical Review* 15(3). 213–225.

Burns, Louis F. 1994. *Symbolic and decorative art among the Osage People*. Fallbrook, CA: Ciga Press.

Carroll, Michael P. 1992. Folklore and psychoanalysis: The Swallowing Monster and Open-Brains allomotifs in Plains Indian Mythology. *Ethos* 20(3). 289–303.

Chafe, Wallace L. 1976. *The Caddoan, Iroquoian, and Siouan languages*. The Hague, Netherlands: Mouton.

Cumberland, Linda A. & Robert L. Rankin. 2012. *Kaáⁿza íe wayáje: An annotated dictionary of Kaw (Kanza), Kaw-English/English-Kaw*. Kaw City, OK: Kaw Nation.

Dahlstrom, Amy. 2003. Warrior powers from an underwater spirit: Cultural and linguistic aspects of an illustrated Meskwaki text. *Anthropological Linguistics* 45(1). 1–56.

Diaz-Granados, Carol, James R. Duncan & F. Kent Reilly, III (eds.). 2015. *Picture cave: Unraveling the mysteries of the Mississippian cosmos*. Austin, TX: University of Texas Press.

Diaz-Granados, Carol, Marvin W. Rowe, Marian Hyman, James R. Duncan & John R. Southon. 2001. AMS radiocarbon dates for charcoal from three Missouri pictographs and their associated iconography. *American Antiquity* 66(3). 481–492.

Dorsey, James Owen. 1894. A study of Siouan cults. In *11th annual report of the Bureau of American Ethnology*, 351–544. Washington, DC: Government Printing Office.

Dorsey, James Owen. n.d. James o. dorsey papers. Ms. (Box 4800 folders 304 & 305. American Anthropological Archives, Smithsonian Institution: Washington, DC).

Duncan, James R. & Carol Diaz-Granados. 2000. Of masks and myths. *Midcontinental Journal of Archaeology* 25(1). 1–26.

Duncan, James R. & Carol Diaz-Granados. 2015. Transmogrification, healing, and resurrection: Extraordinary themes. In Carol Diaz-Granados, James R. Duncan & F. Kent Reilly, III (eds.), *Picture Cave: Unraveling the mysteries of the Mississippian cosmos*, Arlington, TX: University of Texas Press.

Fletcher, Alice C. & Francis La Flesche. 1911. *The Omaha tribe.* Twenty-Seventh Annual Report of the Bureau of American Ethnology. Washington, DC: Government Printing Office.

Foster, Lance Michael. 1994. *Sacred bundles of the Ioway Indians*. Ames, IA: Iowa State University M.A. thesis.

GoodTracks, Jimm. 2008. The Ioway-Otoe-Missouria (Báxoje-Jiwére-Ñutʔačhi) language project and dictionary. Topeka, KS: Kansas Historical Society. (Available at http://www.kshs.org/p/the-ioway-otoe-missouria-baxoje-jiwere-nut-achi-language-project-and-dictionary/16055/).

Greer, Jill. 2019. Hook-swinging giants and other fantastic themes in Jiwere-Baxoje folktales. In Ryan M. Kasak (ed.), *Proceedings of the 38th Siouan and Caddoan Languages Conference (SCLC)*, 45–68.

Hamilton, William. n.d. Extracts from the unpublished life journals of Rev. William Hamilton, a missionary of the Presbyterian Board for 20 years among the Iowas and Sacs of Kansas and the Omaha of Nebraska, now resident at Bellevue, Nebraska. Ms. Lewis Henry Morgan papers, Box 23. Folder 94: 1-34. Rochester, NY: University of Rochester Library.

Hartmann, Iren & Christian Marschke (eds.). 2010. *Hocąk teaching materials vol. 1: Elements of grammar/learner's dictionary*. Albany, NY: SUNY Press.

Helmbrecht, Johannes & Christian Lehmann (eds.). 2010. *Hocąk teaching materials: Elements of grammar/learner's dictionary*. Albany, NY: State University of New York Press.

Howard, James H. 1995. *The Ponca Tribe*. Licoln, NE: University of Nebraska Press.

Jarrell, Jason & Sarah Farmer. 2019. The legends and archaeology of Devil's Lake: A place of ancient power in Wisconsin. (Online article available at https://www.ancient-origins.net/ancient-places-americas/devils-lake-0011896).

Kasak, Ryan M. & Sarah Lundquist. 2019. Nasal harmony in hoocąk and mandan. In Ryan M. Kasak (ed.), *Proceedings of the 38th Siouan and Caddoan Languages Conference (SCLC)*, 99–122.

Kaufman, David. 2014. *The Lower Mississippi Valley as a language area*. Lawrence, KS: University of Nebraska dissertation.

La Flesche, Francis. 1975. *A dictionary of the Osage language* Smithsonian Institution Bureau of American Ethnology Bulletin 109. Phoenix, AZ: Indian Tribal Series. (Originally published in 1932.).

Lewis-Williams, J. David & Thomas A. Dowson. 1988. The signs of all times: Entopic phenomena in Upper Palaeolithic art. *Current Anthropology* 51(200). 173–205.

Marsh, Gordon. n.d. Unpublished manuscripts. Ms. (American Philosophical Society. Philadelphia, PA).

Meeker, Louis L. 1901. Siouan mythological tales. *The Journal of American Folklore* 14(54). 161–164.

Miner, Kenneth L. n.d. A Winnebago lexicon. Ms. (Robert L. Rankin Collection to Missouri Chiwere Language Project. Columbia, MO: University of Missouri-Columbia).

O'Brien, Michael. 1994. *Cat monsters and head pots: The archaeology of Missouri's Pemiscot Bayou*. Columbia, MO: University of Missouri Press.

Peters, Virginia Bergman. 2000. *Women of the earth lodge: Tribal life on the Plains*. Norman, OK: University of Oklahoma Press.

Prentice, Guy. 1986. An analysis of the symbolism expressed by the Birger figurine. *American Antiquity* 51(2). 239–266.

Radin, Paul. 1911. *The ritual and significance of the Winnebago medicine dance*, vol. 24(92).

Radin, Paul. 1948. *Winnebago hero cycles: A study in aboriginal literature.* Indiana University Publications in Anthropology and Linguistics: Memoir 1 of the International Journal of American Linguistics. Bloomington, IN: Indiana University Press.

Radin, Paul. 1950. Winnebago cutlrue as described by themselves. the origin myth of the Medicine Rite: Three versions & the historical origins of the Medicine Rite. In Charles F. Voegelin, Paul Radin, George Herzog & Glenn A. Black (eds.), *Special publications of the Bollingen Foundation No. 2*, 1–78. Baltimore, MD: Waverly Press. (Available at https://archive.org/stream/winnebagoculture00radi/winnebagoculture/radi.divu.txt).

Reilly, F. Kent, III. 2015. The cave and the Beneath World Spirit: Mythic dragons from the North American past. In Carol Diaz-Granados, James R. Duncan & III F. Kent Reilly (eds.), *Picture Cave: Unraveling the mysteries of the Mississippian cosmos*, 134–143. Arlington, TX: University of Texas.

Skinner, Alanson. 1915. Societies of the Iowa. In Clark Wissler (ed.), *Societies of the Plains Indians, anthropological papers of the American Museum of Natural History, vol. XI*, 679–740. New York, NY: The American Museum of Natural History.

Skinner, Alanson. 1925. Tradition of the Iowa Indians. *The Journal of American Folklore* 38(150). 425–506.

Skinner, Alanson. 1926. Ethnology of the Iowa Indians. *Bulletin of the Public Museum: Milwaukee* 5(4). 181–354.

Stanley, Lori A. 1990. *"the Indian path of life: A life history of Truman Washington Dailey*. Columbia, MO: University of Missouri-Columbia.

Stanley, Lori A. 2004. Ratcliffe sacred rock and the seven sacred stones. In Carol Diaz-Granados & James R. Duncan (eds.), *The rock-art of Eastern North America: Capturing images and insight*, 19–41. Tuscaloosa, AL: University of Alabama Press.

Thompson, Stith. 1929. *Tales of the North American Indians*. Cambridge, MA: Harvard University Press.

Thompson, Stith. 1966. *Tales of the North American Indians*. Bloomington, IN: Indiana University Press. (Originally published in 1929).

Whitman, William. 1936. *The Oto*. New York, NY: Columbia University Press.

Unaccusativity in Crow*

Edwin Ko
University of California, Berkeley

Abstract: Like in many other Siouan languages, Crow displays an active-stative (or Split-S) system. This paper explores the semantic features that underlie the active-stative split in Crow and examines a handful of distinct morphosyntactic behaviors between active and stative intransitives. Specifically, the two questions that are addressed in this paper are as follows: What are the semantic factors, if any, that would allow us to predict which class a verb belongs to based on its meaning alone? Is there (morpho)syntactic evidence in Crow to support the unaccusative hypothesis? The overall claim is that unaccusativity in Crow is semantically determined, but syntactically encoded. The semantic base underlying the two verb classes involves an interaction between agentivity and lexical aspect, while several possible unaccusative diagnostics, involving multiple exponence, the "inchoative" morpheme *-dee*, causative alternation and noun incorporation, provide support for an analysis that active intransitives are unergatives and stative intransitives are unaccusatives.

Keywords: Crow, active-stative, split intransitivity, unaccusativity

1. Introduction

Like in many other Siouan languages, Crow displays an active-stative (or Split-S) system.[1] In Crow, this morphosyntactic alignment system is directly observable through first- and second-person agreement markers on intransitive verbs, which belong to one of two classes: ACTIVE or STATIVE. In an active intransitive verb, as in (1), the verb takes the active (A-set) morpheme *baa-* to refer to the first-person subject, which is shown in bold. On the other hand, intransitive verbs that belong in the stative class cross-reference their subjects using stative (B-set) morphemes, such as the first-person morpheme *bii-*, as in (2), which is underlined. The A- and B-set morphemes also appear on active transitive verbs to reference subjects and objects, respectively. Thus, intransitive verbs in Crow take different person morphemes depending on whether they are active or stative: active verbs take A-set markers whereas stative verbs take B-set markers. The full paradigm for A- and B-set morphemes across intransitive verbs are given in Table 1.[2]

*Many thanks to my friends in Crow Country: Felice Big Day, Cyle Old Elk, Jack Real Bird, Riley Singer, Charles Yarlott Jr., and many others for graciously sharing their language and culture with me. I am also grateful to Tyler Lemon, Andrew Garrett, Peter Jenks, Lev Michael, Line Mikkelsen, Zach O'Hagan, Amalia Skilton, Martha Schwarz, and participants at Fieldwork Forum at UC Berkeley and the 39th SCLC for helpful comments on this project. As always, all errors are my own.

[1] See Mithun (1991) for a discussion of the various terms used to refer to an active-stative alignment system.

[2] Abbreviations used in the glosses are as follow: A: A-set, B: B-set, CAUS: direct causative, DECL: declarative, IMPER: imperative, INCHO: inchoative, INDEF: indefinite, INSTR: instrumental, NEG: negation, PL: plural, SS: same-subject marking, 1: first person, 2: second person, 3: third person. Unless otherwise indicated, the data in this paper

(1) Active intransitive verb:
baa*-lisshí-k*
1A-dance-DECL
'I danced' (FBD)

(2) Stative intransitive verb:
bii-wíisshi-k
1B-tell.lie-DECL
'I lied' (RS)

Table 1: A- and B-set agreement morphemes in Crow

	A-SET	B-SET
1SG	*baa-*	*bii-*
2SG	*daa-*	*dii-*
3SG	∅	∅
1PL	*baa-* +PL	*balee-*
2PL	*daa-* +PL	*dii-* +PL
3PL	∅- +PL	∅- +PL

A major challenge for learners of an active-stative language such as Crow is knowing when a verb employs either the A- or B-set morpheme.[3] One obvious approach is for learners to memorize and eventually internalize which class each verb belongs to. With over 900 documented intransitive verbs (see Gordon & Graczyk n.d.), this formidable and daunting task requires significant amounts of dedication, time, and effort. If there are semantic regularities within each of the two classes that learners can instead leverage, then this would certainly be a more effective and less arduous approach. The questions we might then ask are as follow: Are there patterns, based on the kinds of verbs in each class and their meanings, that would allow a learner to reliably predict whether a verb takes the A- or B-set morpheme? Additionally, how else might these patterns help us learn other areas of Crow grammar more effectively?

The goals of this paper are twofold. First, I investigate the semantic factors that determine whether a verb in Crow is likely to be active or stative. What I ultimately claim is that the distinction between active and stative intransitive verbs is indeed based on their lexical meaning. Specifically, I propose that the semantic base underlying the two verb classes involve an interaction between agentivity (i.e. control) and lexical aspect (i.e. states vs. nonstates). The second goal of this paper explores the (morpho)syntactic properties of active and stative verbs in Crow in relation to unergative and unaccusative verbs (i.e. the unaccusative hypothesis), that have been widely discussed in the literature on syntactic theory. Identifying which class a verb is associated with will also allow learners to form sentences in ways that parallel with the language used by fluent speakers of Crow. For example, constructions that involve the so-called modal and continuative auxiliaries, benefactive, imperative, among others, display structural differences that rely on verb class membership. The questions I am concerned with are of a more theoretical nature and consist of investigating not only *what* these differences are, but also *why* these differences exist. Based on four sources of evidence that involve constructions in which contrasting morphosyntactic patterns can be found between active and stative verbs, I analyze active intransitives as unergatives and stative intransitives as unaccusatives.

come from my own fieldwork during 2018 and 2019. Speaker names are abbreviated in the following way: FBD: Felice Big Day, COE: Cyle Old Elk, JRB: Jack Real Bird, RS: Riley Singer, and CY: Charles Yarlott Jr.

[3]This paper sets aside mixed-class verbs that straddle between active and stative classes and that take a mixture of both A and B- morphemes.

This paper is written with three types of readers in mind: general linguists (e.g. specialists in language typology), Crow language educators who have some background in linguistics, and other Siouan linguists and researchers. First of all, this paper provides a description of the semantic classification of active and stative verbs in Crow and focuses on the morphosyntactic differences between these two verb classes. Those seeking to better understand the diversity of active-stative languages may find these descriptions helpful. Second, I provide three proposals for the semantic base of the active-stative alignment in Crow. Crow language educators may find these proposals resourceful since they, in principle, will allow learners to predict the verb's class based on its meaning alone. Third, the methodology employed in this paper is described in detail so that others, especially those in the Siouanist community, may find the approach useful in their own language work. In addition, Siounists engaged in comparative research may also find interest in some of the arguments made in this paper.

2. Semantic factors underlying the active-stative split in Crow

Surveying a number of active-stative languages, Mithun (1991) discusses two primary semantic correlates of the active-stative split. One semantic factor is LEXICAL ASPECT (or AKTIONSART) which concerns the inherent temporal properties of verbs. The second semantic factor involves AGENTIVITY which relates to the degree of agency that the subject has over the situation denoted by the verb. Mithun further distinguishes four subclasses of agentivity: performance, effect, instigation, and control. For example, in her classification of Lakota, she claims that performance, effect, and instigation, but not control are directly involved for verbs that belong to the active set (Mithun 1991:516). That is, in Lakota, active verbs tend to indicate situations that participants perform, effect, and instigate, but not control. Throughout this section, I employ her methodology to test which semantic feature(s) best captures the patterns underlying the active-stative split. While Mithun takes a fine-grained approach towards agentivity, I only focus on the notion of CONTROL, setting aside performance, effect, and instigation from the analysis.

In his unpublished manuscript, Rankin (2004) argues that the semantic factors discussed by Mithun cannot be applied across the Siouan languages. The crux of his argument lies in the exceptions and inconsistencies that surface when attempting to classify verbs based on Mithun's semantic typology. For example, stative verbs like 'fall down' and 'perspire' are counterexamples to a classification based on lexical aspect, while active verbs such as 'snore' and 'sneeze' are problematic for the agentivity account. Yet Rankin's push back to Mithun's classification is based on a comparative, holistic approach to the Siouan language family, rather than on a language-by-language basis. As important as it is to understand how languages within a single language family converge and diverge, a bird's-eye view approach cannot lead to claims that rule out potential semantic factors for individual languages; instead, an in-depth study into each individual language is required.

In terms of the active-stative split in Crow, Graczyk (2007) acknowledges that the division between active and stative verbs is based on their lexical meaning. Specifically, he writes that "it is possible to classify a verb as active or stative simply on the basis of its semantic properties [...] knowledge of the meaning of a verb will almost always enable one to predict its class membership" (Graczyk 2007:177-178). However, Graczyk does not provide what these semantic properties might be, and therefore the goal of this section is to investigate and determine precisely

which features motivate verb class membership in Crow.

In what follows, I focus on intransitive verbs in Crow, for which there are roughly 380 active verbs and 580 stative verbs that appear in Ray Gordon and Randolph Graczyk's unpublished and undated *A Dictionary of Crow*. Most of the data in this section originally come from Gordon & Graczyk's Crow dictionary, but all of the words that appear in this paper were checked with at least one speaker of Crow. Because this paper only deals with intransitive verbs, future investigations into the active-stative split in Crow should also consider (di)transitive verbs and whether the analysis put forth here can be extended to those verbs.

2.1. Fluid-person marking verbs

Fluid-person marking verbs are verbs that can take either A- or B-set morphemes. However, verbs differ in their meaning depending on which morpheme is used. Three examples of fluid-person marking verbs are given in (3–5). As an illustration, let us first consider the verbal stem *daxchí-* in (3). Here, when an A-set morpheme is used, we obtain the meaning 'tie, bind', but when a B-set morpheme is used, we instead get the meaning 'choke (on food), gag'.

(3) *daxchí-* (Graczyk 2007:149):
   a. Active: *daxchí-* 'tie, bind'
   b. Stative: *daxchí-* 'choke (on food), gag'

(4) *xachií-* (Wallace 1993:88; CY):
   a. Active: *xachií-* 'move (location)'
   b. Stative: *xachií-* 'feel movement, being moved'

(5) *daxxálua-* (FBD)
   a. *daxxálua-* 'drag'
   b. *daxxálua-* 'slide'

Fluid-person marking verbs make a particularly fruitful starting point because the shift in meaning has a potential to shed light on the semantic factors underlying the two verb classes. For example, a possible semantic explanation for the difference between 'tie, bind' and 'choke (on food), gag' involves control. That is, the act of tying or binding implies that the subject is in control of their own behavior in relation to the situation at hand, while choking on food or gagging implies a lack of subject control. Similar semantic explanations can be provided for (4) and (5) where A-set marking is associated with subject control while B-set marking corresponds to the lack of subject control.

Another semantic explanation for the difference in meaning between the use of A- and B-set morphemes involves lexical aspect. Lexical aspect, also sometimes referred to as aktionsart, refers to the inherent temporal properties of the verb. Specifically, when the A-set morpheme is used on the verb *xachií-* in (4), we get the meaning 'move (location)', which denotes an event or activity. In contrast, when B-set marking is used with the same verb stem, it now means 'feel movement' or 'being moved', both of which may be construed as states.

In light of the shifts in meaning between the use of A- and B-set morphemes in (3–5), there are two semantic properties that we should consider. First, as in (3–5), the shifts seem to be one of control; that is, A-set verbs denote events that are controlled by the subject, whereas B-set verbs

denote events that are not controlled by the subject. Second, as in (4), the shift could also be one of lexical aspect; A-set verbs denote events, whereas B-set verbs denote states. Based on these two semantic factors, we will now examine how well these predictors fare in accounting for verbs in the active and stative class.

2.2. Proposal #1: Lexical aspect

To what extent does lexical aspect explain which verbs fall in either the active or stative class? As their names suggest, one might expect active verbs to denote types of actions, while stative verbs denote states. Using the basic aspectual distinction STATE and EVENT (or NONSTATE), in which the latter involves a change to the situation, whereas the former does not (Dowty 1979), I group each verb based on the following criteria: Is the verb active or stative? Does the verb denote an event or a state? From this grouping, we obtain four discrete categories of possible verb types: (i) active verbs that denote events, (ii) stative verbs that denote states, (iii) active verbs that denote states, and (iv) stative verbs that denote events. The former two follow the prediction that active verbs denote events and stative verbs denote states, while the latter two are exceptions and present counterexamples to the proposal.

Overall, lexical aspect appears to be a reliable predictor of active and stative intransitive verbs – most active verbs do in fact denote events and most stative verbs do indeed denote states, with examples provided in (6) and (7). In other words, verbs denoting events, such as *díili-* 'walk' and *xalússhi-* 'run', tend to be classified as active verbs that take A-set morphemes while verbs denoting states, such as *chía-* 'be white' and *isáa-* 'be big', are likewise often classified as stative verbs that take B-set morphemes.

(6) Active intransitives denoting events:

a. *díili-* 'walk'
b. *xalússhi-* 'run'
c. *iluú-* 'stand up'
d. *ilíi-* 'talk, speak'
e. *baachimmí-* 'study'
f. *disshí-* 'dance'
g. *chiwakíi-* 'pray'
h. *káa-* 'laugh, smile'
i. *shée-* 'die, faint'
j. *bilihpí-* 'bathe, swim'

(7) Stative intransitives denoting states:

a. *chía-* 'be white'
b. *isáa-* 'be big'
c. *satchí-* 'be thick'
d. *shishía-* 'be dirty'
e. *tawée-* 'be hot'
f. *ítchi-* 'be good'
g. *púmmi-* 'be short'
h. *xaliá-* 'be itchy'
i. *xawíi-* 'be bad'
j. *xusshí-* 'be fast'

Even though there is a strong tendency for active and stative intransitive verbs to denote events and states, respectively, there are still some exceptions, which are shown in (8) and (9). Verbs in (8), such as *chilíi-* 'be afraid', denote states but pattern like other active verbs in taking A-set morphemes, while verbs in (9) denote events and take B-set morphemes. The exceptions, while more numerous in the stative class, are counterexamples for the proposal that lexical aspect is the

single semantic property governing class membership. In order to assess the overall effectiveness of this proposal, we must also consider the reliability of other possible semantic factors. Only then will we able to compare the effectiveness of every proposal and determine the most appropriate semantic predictor of class membership in Crow.

(8) Active intransitives denoting states:

a. *chilíi-* 'be afraid'
b. *ilutchítchi-* 'feel guilty'

(9) Stative intransitives denoting events:

a. *bíile-* 'tell on, tattle'
b. *bíisshi-* 'tell a lie'
c. *ámmichi-* 'fall down'[4]
d. *apáali-* 'grow, sprout'
e. *chipí-* 'drown'
f. *daxxálua-* 'slide, skid'

2.3. Proposal #2: Agentivity

The second proposal involves the notion of AGENTIVITY. Specifically, as we have gleaned from the shifts in meaning across the fluid-person marking verbs in §2.1, another possible semantic base behind the active-stative split is CONTROL; that is, does the subject have control over their own behavior with regards to the situation denoted by the verb? In the literature, control is sometimes referred to by the broader term agentivity, and although not without problems, I use those two terms interchangeably in this paper.[5]

As with lexical aspect, I classify verbs using the following criteria: Is the verb active or stative? Does the subject of the intransitive verb have control over the situation denoted by the verb? While control may be viewed as a gradient phenomenon, for the purpose of this study, I treat it as categorical: either the subject has some control of the situation or the subject does not. The resulting categories are as follows: (i) active verbs with subjects that have control over the situation, (ii) stative verbs with subjects that lack control of the situation, (iii) active verbs with subjects that lack control of the situation, and (iv) stative verbs with subjects that have control over the situation. The former two categories are in line with the prediction that active verbs consist of subjects that control, to some degree, the situation, while stative verbs have subjects that completely lack control. Again, the latter two groups are exceptions and therefore should be subject to further scrutiny.

What we find is that agentivity is also a relatively reliable predictor of verb class membership – most active intransitives do indeed have subjects that control the situation, as in (10), whereas stative intransitives have subjects that lack control, as in (11). For example, verbs in (10), such as *disshí-* 'dance' and *chiwakíi-* 'pray', involve participants that are in control of their own dancing and praying. In contrast, verbs in (11), such as *isáa-* 'be big' or *púmmi-* 'be short', consist of subjects that do not control the situation – in general. one does not have control over their own inherent physical attributes.

[4] *ámmichi* > *awé* 'ground' + *dichí* 'hit' (cf. *awélichi-* 'fall down').

[5] See Duranti (2004) for a detailed discussion of agency in language and the proliferation of different usages across other disciplines.

(10) Active intransitives denoting events:

a. *díili-* 'walk'
b. *xalússhi-* 'run'
c. *iluú-* 'stand up'
d. *ilíi-* 'talk, speak'
e. *baachimmí-* 'study'
f. *disshí-* 'dance'
g. *chiwakíi-* 'pray'
h. *bilihpí-* 'bathe, swim'
i. *iiwaannía-* 'play'
j. *dée-* 'go'

(11) Stative intransitives denoting states:

a. *chía-* 'be white'
b. *shipíta-* 'be black'
c. *isáa-* 'be big'
d. *ítchi-* 'be good'
e. *púmmi-* 'be short'
f. *háchka-* 'be tall, long'
g. *satchí-* 'be thick'
h. *xaliá-* 'be itchy'
i. *xawíi-* 'be bad'
j. *chilía-* 'be cold'

Although the results so far suggest that active verbs tend to involve participant control, while most stative verbs do not, classifying verbs based on agentivity reveals a number of exceptions. These exceptions include active intransitives where participants lack control and stative intransitives where participants have some control over the situation. The counterexamples, which are more numerous in the active class, are displayed in (12) and (13), respectively.

(12) Active intransitives with participants lacking control of the situation:

a. *axxí-* 'cough'
b. *apiiaxxí-* 'sneeze'[6]
c. *pía-* 'fart, break wind'
d. *kalée-* 'vomit'
e. *chilíi-* 'be afraid'[7]
f. *shée-* 'die, faint'
g. *baashíali-* 'dream'
h. *páxpi-* 'suffer, feel pain'
i. *ilutchítchi-* 'feel guilty'
j. *iháwi-* 'sleep'

(13) Stative intransitives with participants controlling the situation:

a. *bíile-* 'tattle'
b. *bíisshi-* 'tell lie'

Among the exceptional verbs in the active class shown in (12) are verbs denoting internally-caused bodily processes, such as 'cough', 'fart' and 'vomit'. Despite the fact that these kinds of events are not directly caused by the participant, cross-linguistically these verbs form a homogeneous subset of verbs that pattern in some ways like agentive (or active) intransitive verbs (Levin & Hovav 1995).[8] If verbs denoting internally-caused bodily processes are typically classified in the same way as verbs where participants have direct control over the situation, then perhaps we can

[6] *apiiaxxí-* > *apá* 'nose' + *ii* 'instrumental' + *axxí* 'cough' (Gordon & Graczyk n.d.).
[7] Cf. *chilíihche* 'forbid someone to do something' > *chilíi-* 'be afraid' *-hche* 'indirect causative'.
[8] I speculate that the reason why internally-caused bodily processes tend to be classified as active verbs is that they involve indirect causation by *animate* participants.

set aside these so-called "exceptions." In doing so, the number of true exceptions for active verbs (e.g. *chilíi-* 'be afraid', *ilutchítchi-* 'feel guilty') dwindles significantly, and we can conclude that agentivity is a more reliable predictor than lexical aspect since it introduces less exceptions across the two classes overall.

If verbs that express internally-caused bodily processes pattern like active verbs, then we should not expect to find any such verbs in the stative class. Yet there is a handful of verbs that in fact express internally- caused bodily processes belonging to the set of stative verbs, given in (14). For example, *tannáa-* 'shiver' and *ilítshia-* 'stick, be smelly' are verbs that are found among other verbs in the agentive, active class in other languages, such as Italian.

(14) Stative intransitives internally-caused bodily processes:

a. *táwasaali-* 'sweat, perspire'[9]
b. *bixúa-* 'have diarrhea'
c. *alíishxachii-* 'shake due to hunger'[10]
d. *íishpuuxachii-* 'have cramps in stomach'[11]
e. *apíiluu-* 'have nosebleed'[12]
f. *shéhchikiichi-* 'hiccough'
g. *tannáa-* 'shiver'
h. *ilítshia-* 'stink, be smelly'
i. *daxchí-* 'choke (on food), gag'

How might we try to explain why verbs expressing bodily processes appear in both verb classes? One analysis involves varying degrees of participant control. That is, the more control a subject has over their own bodily process, the more likely it will be classified as an active verb, and the less control a subject has over their bodily processes, the more likely it is classified as a stative verb. For example, participants may exert more control over their flatulence (i.e. *pía-* 'fart, break wind') than their perspiration (i.e. *táwasaali-* 'perspire, sweat') and consequently, the former falls in the active class while the latter is in the stative class. But does the subject of the active verb *kalée-* 'vomit' have more control over their vomiting than, say, their own diarrhea with the stative verb *bixúa* 'have diarrhea'? In these cases, it is not so clear and may depend on various variables. Therefore, there is an inherent difficulty in determining which processes are more or less controlled by the subject, especially when we consider the possibility of cultural variation or culture specific patterns, where the degree of subject control may vary across different cultures. I leave the delineation between the two sets of bodily processes to future investigation but caution those who wish to take up this challenge to exercise utmost care.

While lexical aspect and agentivity both admit some exceptions, I claim that agentivity is a more reliable predictor. When lexical aspect is the main semantic factor, more exceptions are found in the stative class. In diametric opposition, with agentivity as a semantic base, more exceptions appear in the active set. This difference in size between the two sets of exceptions indicates that lexical aspect accounts for more active verbs while agentivity accounts for more stative verbs. However, the number of exceptions is far greater under the proposal involving lexical aspect than agentivity and consequently, I argue agentivity wins over lexical aspect. In the next section, I explore an additional proposal that combines lexical aspect and agentivity.

[9] *táwasaali-* > *tawée* 'hot' + *asaalí* 'come out, exit'.
[10] *alíishxachii-* > *alíishi* 'be hungry' + *xachíi* 'move'.
[11] *íishpuuxachii-* > *íishpuu* 'stomach' + *xachíi* 'move'.
[12] *apíiluu-* > *apá* 'nose' + *ilúu* 'persist'.

2.4. Proposal #3: Lexical aspect and agentivity

The third proposal considers both lexical aspect and agentivity in predicting verb class membership. The hypothesis that I investigate here is stated in the following way: The split between active and stative verbs is based on an interaction between lexical aspect and agentivity—active verbs denote events that participants have control over, while stative verbs denote states that lack participant control. Because both lexical aspect and agentivity on their own are able to predict the majority of verbs, albeit with some exceptions, it follows that the combination of lexical aspect and agentivity should also provide a reliable diagnostic for determining which verbs belong in the active or stative set. Therefore, I focus on analyzing the exceptions that result from proposing a semantic factor involving an interaction between lexical aspect and agentivity as they provide greater insights into the effectiveness of the proposal. The exceptions found in the active set (i.e. uncontrolled states) and stative set (i.e. controlled nonstates) are displayed in (15) and (16), respectively.

(15) Active intransitives that denote states lacking participant control:

a. *chilíi-* 'be afraid'
b. *páxpi-* 'suffer, feel pain'
c. *ilutchítchi-* 'feel guilty'

(16) Stative intransitives denoting events with participant control:

a. *bíile-* 'tell on, tattle'
b. *bíisshi-* 'tell a lie'

The verbs in (15) and (16) are what I consider as true exceptions to the current investigation. In other words, I do not consider active intransitives that denote events but are not controlled by the participant (e.g. *shée-* 'die, faint', *axxí-* 'cough') or stative verbs that denote events but lack participant control (e.g. *ámmichi-* 'fall down', *apáali-* 'grow, sprout') as exceptions. Table 2 displays the distribution of Crow active and stative intransitive verbs based on the interaction between agentivity and lexical aspect.

Table 2: Distribution of Crow active and stative verbs based on an interaction between agentivity (i.e. control vs. non-control) and lexical aspect (i.e. event vs. state). Cells with a darker shade indicate a higher frequency of verbs for that particular interaction, and cells with lighter shade indicate lower frequency.

		EVENT	STATE
CONTROL	ACTIVE	*run, dance*	*sit, live*
	STATIVE	*tattle, tell a lie*	*be cruel, be lazy*
NON-CONTROL	ACTIVE	*die, vomit*	*be afraid, suffer*
	STATIVE	*slip, fall*	*be tall, be big*

Most verbs in Crow tend to cluster in two significant ways. First, there is a very strong tendency for verbs that denote events with participant control to be classified as active verbs. Second, verbs that indicate states with participants lacking control tend to be categorized as stative verbs. The next most frequent cells are verbs that denote events without participant control and these verbs can be found in both active and stative sets. Recall that many of the active and stative

verbs that belong to this category are exceptions found where the semantic factor was either agentivity or lexical aspect (see (9) and (12)). Less frequent are verbs that denote (semi-)states where participants have some degree of control and these verbs appear in both the active and stative class. Finally, verbs in (15) and (16), which I consider true exceptions to the present proposal, are the least frequent in their respective categories: active verbs that denote states lacking participant control and stative verbs that denote events with participant control. The distribution of verbs ranked by their frequency is given in Table 3.

Table 3: Distribution of Crow active and stative in order of most frequent (left) to least frequent (right) based on an interaction between agentivity (control vs. non-control) and lexical aspect (event vs. state).

	MOST FREQUENT			**LEAST FREQUENT**
ACTIVE	[EVENT, CONTROL]	[EVENT, NON-CONTROL]	[STATE, CONTROL]	[STATE, NON-CONTROL]
STATIVE	[STATE, NON-CONTROL]			[EVENT, CONTROL]

But why should verbs that denote events and lack participant control outrank verbs that denote states with participant control? While it is not clear to me exactly why this should be, it is interesting to note that states tend to strongly imply lack of control on the part of the subject. For example, as mentioned in §2.3, states, especially inherent attributes such as being big or being short, are typically not controlled by the subject. This observation is crucial also because it points to the intimate relationship between lexical aspect and agentivity. Still, active and stative verbs that denote uncontrolled states and controlled events, respectively, remain the lowest ranked categories in terms of their overall frequency.

An investigation into the proposal that an interaction between lexical aspect and agentivity underlies the active-stative split reveals that the classification of verbs in Crow does not have to rely on a single semantic criterion. Rather, lexical aspect and agentivity overlaps in complex but regular ways and verbs may fall at different points in the spectrum of this interaction, constrained by what is conceptually possible in language. Compared to the previous two semantic factors, this proposal has the most empirical coverage since it is able to account for most intransitive verbs in Crow and introduce the least number of exceptions. Additionally, this account fits squarely into the typological space of languages that also display an interaction between agentivity and lexical aspect, such as Nepali (Li 2007) and other Indo-Aryan languages (Schwarz, p.c., 2019). Moreover, it aligns with the notion that the semantic base underlying active-stative splits can shift over time (see Mithun 1991, Pustet 2002, and Rankin 2004)–the close interaction or relationship between the dual semantic bases provides one way in which this change can occur.

2.5. Interim conclusion

We have considered three proposals that allowed us to predict which verbs fall into which class based solely on their meaning: lexical aspect (ASPECT), agentivity (AGENT), and an interaction between lexical aspect and agentivity (ASPECT+AGENT). We observed that all three proposals are able to account for the majority of intransitive verbs in Crow and that all three are prone to some number of exceptions. In particular, proposal ASPECT+AGENT introduces the least number of exceptions, and consequently, of the three proposals, it has more explanatory power and empirical coverage.

Yet proposal ASPECT+AGENT requires an additional semantic factor to be stipulated. Even more so, it does not provide an account for verbs that straddle between the leftmost and rightmost (i.e. most and least frequent) columns in Table 3. For example, this proposal does not help us determine whether verbs that denote uncontrolled events belong to the active or stative class. From the perspective of a learner of Crow, this point is critical.

What then is the best tool that we can provide learners to effectively generalize across most intransitive verbs in Crow? Proposal AGENT has an advantage over proposal ASPECT—the former accounts for more verbs and leads to less exceptions. As we have discussed, many of the exceptions of proposal AGENT also form a coherent set of bodily processes (see (12)). If learners simply memorize these and other exceptions, then using agentivity as an active-stative diagnostic will cover most, if not all, of the intransitive verbs. Thus, proposal AGENT is the best tool we have to provide students to empower their own learning of Crow. In other words, *active verbs tend to involve subjects that have control over their own behavior in relation to the situation*.

From an analytic standpoint, I ultimately claim that proposal ASPECT+AGENT best captures the patterns of the active-stative split across intransitive verbs. First, the natural relationship between agentivity and lexical aspect results in the broadest coverage of intransitive verbs in Crow. Second, as we will see in §3.2, the exceptions of proposal ASPECT+AGENT are subject to variation across speakers suggesting a possible change in progress; exceptional verbs involving bodily functions under the proposal AGENT are not subject to such variation. Finally, although proposal AGENT has better direct applications in language teaching, one still has to grapple with the exceptions that involve internally-caused bodily functions within the active set.

3. Unaccusative diagnostics

In the previous section, I have suggested three possible semantic proposals to account for intransitive verbs in their respective classes. Each proposal is able to explain the majority of intransitives to varying degrees and whatever is the optimal proposal is up for debate. Nevertheless, it is clear that verb class membership in Crow is semantically determined; that is, for a given intransitive verb, if one knows its meaning, then one can reliably predict which verb class it belongs to. In this section, I explore the differences in morphosyntactic behaviors between active and stative intransitive verbs to show that while the two verb classes are semantically predictable, they also represented syntactically in different ways. These different syntactic encodings thus explain why we observe such different structural behaviors and understanding which class a verb belongs to will allow learners to construct sentences that fit the structural profile of other verbs within the same set.

Before we proceed, it is important to briefly overview the UNACCUSATIVE HYPOTHESIS. In his seminal paper, Perlmutter (1978) proposes that intransitive verbs consist of two classes, namely UNERGATIVES and UNACCUSATIVES, and that each of these two classes of verbs corresponds to a distinct syntactic configuration. Specifically, unergative verbs introduce subjects (external arguments) that semantically and syntactically behave like subjects of transitive verbs, while unaccusative verbs introduce subjects (internal arguments) that semantically and syntactically behave like objects of transitive verbs.[13] For example, like subjects of many transitive verbs, subjects of unergatives tend to be agentive while, in contrast, much like objects of transitive verbs, subjects of unaccusatives

[13] While Crow has both active transitives and stative transitive, in this paper I use the term transitive to refer only to the set of active transitives.

tend to be non-agentive; that is, these arguments tend to undergo some action that is controlled by some external force or entity.

The unaccusative hypothesis has been thoroughly tested on languages from a diverse set of alignment systems (e.g., see Levin & Hovav 1995 and references therein). For a variety of active-stative languages, many scholars have analyzed active intransitives as unergative verbs and stative intransitives as unaccusative verbs (e.g., Rice 1991 for Slave, Williamson 1979 and Legendre & Rood 1992 for Lakota, Wallace 1993 for Crow, West 2003 for Nakota, Boyle 2007 for Hidatsa, *inter alia*). However, it is still not clear where Crow fits into this proposal. Wallace (1993:94-99) presents two main arguments for unaccusativity in Crow that rely on (a) the semantic characteristic of intransitive verbs and (b) its agreement morphology. It turns out, according to Levin & Hovav (1995), that these kinds of arguments are insufficient evidence for the unaccusative hypothesis. Ideally, one would provide arguments based primarily on syntactic evidence.

Is there syntactic evidence for a one-to-one correspondence in Crow between active intransitives and unergatives, and stative intransitives and unaccusatives? To address this question, I examine a number of differences in the morphosyntactic behaviors between the two sets of intransitives and determine to what extent they are able to support the unaccusative hypothesis. Ultimately, I identify four possible diagnostics: multiple exponence, the so-called "inchoative" morpheme *-dee*, causative alternation, and noun incorporation, as support for the existence of an unaccusativity split in Crow. Based on these four sources of evidence, I analyze active intransitives as unergatives and stative intransitives as unaccusatives.

3.1. Unaccusative diagnostic #1: Multiple exponence

Multiple exponence is the multiple realization of a single morphological feature within a word (Harris 2017). In Crow, there is a handful of constructions where multiple exponence obligatorily occurs. These constructions involve so-called modal auxiliaries (i.e. *-bia* 'want/going to', *-iimmaachi* 'will', etc.), continuative auxiliaries (i.e. *-daachi* 'keep on (continuous)', *-dahku* 'keep on (iterative)', etc.), and the benefactive marker *-ku*. Across these constructions, there is a divide between active and stative intransitive verbs in the realization of multiple exponence. For example, with modal auxiliary *-bia* 'going to', multiple exponence occurs only for active intransitives but not for stative intransitives. These two morphosyntactic behaviors, with multiple exponence given in bold, are displayed in (17) for active intransitive *disshí-* 'dance' and (18) for stative intransitive *ámmichi-* 'fall'. Moreover, multiple exponence occurs with active transitive verbs, such as *dichí-* 'hit' in (19).

(17) a. *baa-lisshí-k*
1A-dance-DECL
'I danced' (CY)

b. ***baa**-lisshí-wia-**waa**-k*
1A-dance-going.to-1A-DECL
'I'm going to dance' (RS)

(18) a. *bii-ámmichi-k*
1A-fall-DECL
'I fell' (CY)

b. *bii-ámmit-bia-k*
1A-dance-going.to-1A-DECL
'I'm going to fall' (JRB)

(19) a. *dii-waa-lichí-k*
2B-1A-hit-DECL
'I hit you' (CY)

b. *dii-**waa**-(l)ít-bia-**waa**-k*
2B-1A-hit-going.to-1A-DECL
'I'm going to hit you' (JRB)

Notably, the morpheme that is being multiply exponed is the first-person A-set marker that cross-references the agentive subject in the verbs *disshí-* 'dance' and *dichí-* 'hit'; the B-set marker that refers to non-agentive subjects and objects fails to exhibit multiple exponence, as with the verb *ámmichi-* 'fall'. Under the assumption that multiple exponence in Crow is syntactically motivated, the fact that the A-set agreement morpheme targets subjects of active intransitives and of transitive verbs and is subject to multiple exponence demonstrates that these arguments undergo similar syntactic processes and share similar syntactic configurations.[14] This point is crucial because the parallel behavior between subjects of active intransitive verbs *and* transitive verbs provides direct support for the unaccusative hypothesis. In contrast, B-set proclitics only refer to subjects of stative intransitives and objects of transitive verbs, but they may not be exponed more than once. Therefore, the claim is that an intransitive verb that realizes multiple exponence is unergative while a verb that displays no multiple exponence is unaccusative.

3.2. Unaccusative diagnostic #2: Inchoative *-dee*

The second unaccusative diagnostic involves the so-called "inchoative" suffix *-dee* which is roughly translated as 'become'. This suffix, which combines only with stative and not active intransitives, appears in imperative and desiderative constructions, and conjugates irregularly for the person of the subject. Below, I first describe the distribution of the inchoative in imperatives and then in desiderative constructions.

The imperative suffix *-ah* is used to express commands or requests. The imperative suffix attaches directly to active intransitive verbs (20) as well as transitive verbs (21). However, this suffix may not attach directly to stative verb stems. To form imperatives with stative intransitive verbs, the inchoative suffix *-dee* (underlined) must also be used, as in (22).[15]

(20) a. *Baláxi-h!*
sing-IMPER
'Sing!' (Graczyk 2007:151)

b. *Disshí-h!*
dance-IMPER
'Dance!' (COE)

(21) a. *Duushí-h!*
eat-IMPER
'Eat it!' (COE)

b. *Dichí-h!*
hit-IMPER
'Hit it!' (Graczyk 2007:151)

[14] While the nature of the agreement system of Crow is beyond the scope of this paper, interested readers are referred to Ko (2019) for an analysis of Crow as a split-agreement system consisting of φ-agreement and clitic doubling (or pronominal cliticization).

[15] The imperative suffix in Crow triggers ablaut that is a key feature of Siouan languages. For example, morpheme-final *-ee* is realized as *-aa* when directly preceding the imperative, as can be seen in (22) with the inchoative *-dee*.

(22) a. *Ítchi-laa-h!*
be.good-INCHO-IMPER
'Be good!' (FBD)

b. *Ámmit-daa-h!*
fall-INCHO-IMPER
'Fall!' (FBD)

The use of the inchoative highlights the dichotomous relationship between active and stative intransitives in the formation of imperatives. Moreover, it shows the analogous behavior between active intransitives and transitive verbs. With active intransitives and transitives, the imperative suffix directly attaches to the verbal stem, but with stative intransitives, the imperative suffix attaches to the derived stem with the inchoative.

But why should such a split exist between active and stative intransitive verbs? One line of analysis is to suggest that imperatives cannot combine with verbs denoting states. In fact, Lakoff (1966) proposes that imperatives are a test for stativity. Under this approach, stative intransitives can only be used in the imperative when combined with the inchoative—in theory, the inchoative converts stative intransitives into verbs denoting a change of state or achievement (i.e. *to be S* → *to become S*, where S is a state). More recently, however, there have been scholars who argue that imperatives are actually a test for agentivity, and not stativity (Dowty 1979:112, Levin & Hovav 1995:170-171, Jackson 2005). I follow these scholars in assuming that the imperative is not a test of stativity but of agentivity.

So what light does imperative formation shed on the syntactic encoding of active and stative intransitives? Speculating on the function of the inchoative, Wallace (1993:139) writes, "I often think of [*-dee*] as 'volitional be/become', since (in my analysis) it creates an unergative predicate from an unaccusative predicate." Besides appearing with the imperative, the inchoative also appears in desiderative constructions involving stative intransitives, as in (23). Note that the inchoative conjugates for first person in these examples. What is particularly striking is across these desiderative constructions, multiple exponence (bolded), which is generally restricted to the class of active intransitives and transitives, appears adjacent to the desiderative marker *-bia*.

(23) a. *háchka-wee-wia-waa-k*
tall-1A.INCHO-DESID-1A-DECL
'I want to be tall' (JRB)

b. *ámmit-**bee**-wia-**waa**-k*
fall-1A-INCHO-DESID-1A-DECL
'I want to fall' (JRB)

I suggest that the most appropriate analysis is that the inchoative does indeed derive an unergative verb from a stative, unaccusative verb, as first suggested by Wallace. This analysis has the following outcomes. First of all, the imperative suffix *-ah* requires the verb to have an agentive subject. Since subjects of unergatives are generally agentive, this fits neatly within the claims that imperatives are a test for agentivity. Therefore, the only way to make imperatives of stative verbs is to combine them with the inchoative marker *-dee* which transforms them from unaccusatives into unergatives and imbues the argument with agentive properties. Evidence for this transformation comes from desiderative constructions with stative intransitives which realizes multiple exponence, as in (23). These constructions also necessarily involve the inchoative suffix which is unsurprising since the desiderative generally implies a volitional subject. Since multiple exponence typically only occurs with unergative verbs, this provides further support for analyzing stative verb with inchoatives as derived unergatives.

While the intransitive verbs given in (20), (22), and (23) above represent clear cases of active and stative verbs that may or may not take the inchoative, there is a handful of verbs that

optionally allow the inchoative in the imperative and desiderative constructions. For example, the verbs in (24) optionally take the inchoative *-dee*. These verbs belong to the set of exceptional verbs found in §2 for proposals involving agentivity and an interaction between lexical aspect and agentivity. However, active verbs expressing internally-caused bodily processes are not subject to the same kind of optionality—these verbs do not permit the use of the inchoative.

(24) a. *Chilíi(-lee)-ssaa-h!*
be.afraid(-INCHO)-NEG-IMPER
'Don't be afraid!' (FBD)

b. *Bíile(-lee)-ssaa-h!*
tattle(-INCHO)-NEG-IMPER
'Don't tattle!' (FBD)

c. *Páxpi(-lee)-ssaa-h!*
suffer(-INCHO)-NEG-IMPER
'Don't suffer!' (FBD)

d. *Bíisshi(-lee)-ssaa-h!*
tell.lie(-INCHO)-NEG-IMPER
'Don't lie!' (FBD)

According to my consultants, whether an inchoative is used for the exceptional verbs in its imperative form is subject to intra- and inter-speaker variation.[16] The optional use of the inchoative also occurs in desiderative constructions, for example, with the verb *bíisshi-* 'tell a lie', as in (25) and (26). In (25), which consists of a third person subject, the inchoative may be optionally expressed. Since third person marking in Crow is null, examples consisting of a first-person subject is given in (26). In (26a), without the inchoative, the constructions patterns like other desiderative constructions with canonical stative intransitives in lacking multiple exponence. In contrast, multiple exponence occurs when the inchoative is used, as in (26b).

(25) *Logan bíishi(-lee)-wia-k*
Logan tell.lie(-INCHO)-DESID-DECL
'Logan wants to lie' (FBD)

(26) a. *bii-wíisshi-wia-k*
1B-tell.lie-DESID-DECL
'I want to lie' (FBD)

b. *bii-wíisshi-**wee**-wia-**waa**-k*[17]
1B-tell.lie-1A.INCHO-DESID-1A-DECL
'I want to lie' (FBD)

What this optionality suggests is that these exceptional verbs may in fact straddle both active and stative classes in their syntactic representations. With the inchoative, these verbs behave like any other unaccusative verb since unaccusative intransitives require the inchoative to co-occur in imperative and desiderative constructions. On the other hand, these same exceptional verbs may also leave out the inchoative entirely, which is a property of active verbs. Yet these verbs are not quite fully unergatives. In (26a), in which the inchoative is absent, multiple exponence still does not manifest as we would expect from a true active, unergative verb. I suggest that these verbs are undergoing a shift from one class to another. That is, a verb such as *bíishi-* may be in the process of being reanalyzed as an active verb rather than a stative verb, which is unsurprising given their exceptional status. If verbs are semantically determined and syntactically represented in a certain

[16] This pattern can also be found for the set of so-called mixed-class verbs, which draw from both A- and B-set morphemes, in imperative and desiderative constructions (see also Wallace (1993:142-143).

[17] This sentence is an outlier in being the only example in my corpus where A- and B-set cross-references the same argument; in all other cases, A- and B-set only refers to distinct arguments.

way, then a verb like *bíishi-* which patterns semantically like an active verb would be prone to being reanalyzed syntactically as an active intransitive verb over time by some speakers.

To sum up, the inchoative morpheme *dee-*, which occurs in imperative and desiderative constructions, attaches only to unaccusative verbs, not unergative or transitive verbs. Furthermore, the inchoative has an additional function of transforming unaccusatives into unergative verbs thereby allowing multiple exponence to occur.

3.3. Unaccusative diagnostic #3: Causative alternation

The direct causative in Crow *-ee* is a valence-increasing operation that introduces an additional argument, typically a causer or an agent, into the clause, and conjugates for the person of the subject. In other words, it transforms intransitive verbs into transitive ones. Although the direct causative is not particularly productive in Crow, the application of this morpheme is almost always restricted to stative verbs regardless of whether they denote states or nonstates, as in (27) and (28), where (i) represents stative verb stems and (ii) represents stative verb stems with the direct causative.[18]

(27) a. i. *óoshi-* 'be ripe, cooked'
ii. *óosshee-* 'cook'
b. i. *koowí-* 'be complete'
ii. *koowée-* 'complete, quit'
c. i. *úuchi-* 'be dry'
ii. *úutchee-* 'dry (something)'

(28) a. i. *apáali-* 'grow, sprout'
ii. *apáalee-* 'raise'
b. i. *passhí-* 'fall off'
ii. *passhée-* 'make fall down'
c. i. *xapí-* 'fall'
ii. *xapée-* 'drop'

The restriction on the direct causative suggests that the direct causative is sensitive to the syntactic configuration of the verb. In particular, if the verb already has an agentive argument in its syntactic representation (or argument structure), such as active intransitive verbs—which generally have agentive subjects—then the direct causative is blocked from attaching to the verb. In fact, there is evidence that the argument the direct causative introduces behaves like subjects of active intransitive and transitive verbs. First, in the imperative of the stative verb *úuchi-* 'be dry', the inchoative (underlined) must be used, as shown in (29a). With the direct causative on the same verb stem, however, the inchoative is no longer needed, as in (29b), when giving a request or command for something to be dried. Second, among constructions that trigger multiple exponence, such as the modal auxiliary *-bia* 'going to', multiple exponence is absent when the stative verb appears with just the modal auxiliary, as in (30a). In the same construction with a direct causative given in (30b), however, multiple exponence (bolded) occurs on the desiderative morpheme; the direct causative here conjugates for first person.

(29) a. *Úut-<u>daa</u>-h!*
be.dry-INCHO-IMPER
'Be dry!' (FBD)

b. *Úutt-ah!*[19]
dry-IMPER
'Dry it!' (FBD)

[18] One notable exception involves the set of motion verbs that are derived from the form *xii*. These motion verbs behave like active verbs but may take direct causatives.

[19] In Crow, *ch* occurs as *t* before low vowels and obstruents.

(30) a. *bii-úut-bia-k*
1B-be.dry-DESID-DECL
'I'm going to be dry' (FBD)

b. *baaaxúassee úutt-**baa**-wia-**waa**-k*
clothes dry-1A.CAUS-DESID-1A-DECL
'I'm going to dry the clothes' (FBD)

The overall claim here is that the newly introduced subjective argument of a stative verb that has been combined with the direct causative behaves syntactically and semantically like other subjects of active intransitive and transitive verbs. In light of this, direct causatives can only attach to unaccusative verbs because these verbs lack an agentive (external) argument; the direct causative introduces an external argument so if the verb already subcategorizes for one, the direct causative is prevented from applying altogether.

3.4. Unaccusative diagnostic #4: Noun incorporation

In Crow, noun incorporation is a syntactic process whereby a noun combines with and becomes incorporated into the verb Graczyk (2007:7, 293-297).[20] Noun incorporation, which is given in brackets, is attested only for nouns that are objects of transitive verbs, as in (31), and nouns that are subjects of stative intransitive verbs, as in (32). Active verbs, on the other hand, do not allow incorporation of their subjects and attempts to elicit compounding of nouns that are subjects of active intransitive and transitive verbs have so far been unsuccessful. These two sets of behaviors indicate that subjects of stative intransitive verbs and objects of transitive verbs once again pattern in syntactically similar ways thereby providing further support for the claim that stative verbs behave like unaccusative verbs.

(31) a. *Hinné baapé* [*Apsáalook-ilaa*]*-u ii baa-waachimmí-k*
this day Crow-talk-PL INSTR 1A-study-DECL
'Today I learned to speak Crow' (FBD)

b. *Logan* [*bishka-lúupia*]*-k*
Logan dog-dislike-DECL
'Logan dislikes dogs' (FBD)

(32) a. [*ilúk-hilahp*]*-ak*
meat-scarce-SS
'meat is scarce' (Graczyk 2007:282, Ex.21)

b. [*balás-itchi*]*-k*[21]
my.heart-be.good-DECL
'I feel good (lit. my heart is good)' (JRB)

[20] See Golston, Boyle & Gebhardt (2018) for a view that incorporation in Crow is purely phonological rather than syntactic. However, if noun incorporation is strictly phonological, as they claim, there is a question as to how phonology can successfully capture the unaccusativity patterns discussed here.

[21] Although *balásitchi-* 'be happy' is a lexicalized stative verb, its inclusion is to show that at one point in its diachrony, the subject *balás* 'my heart' of the stative verb *ítchi-* 'be good' has been incorporated into the verb.

4. Conclusions

In this paper, I have presented three proposals for the semantic base underlying the active-stative split in Crow. Although all three proposals are able to account for most intransitive verbs in Crow, there are still some exceptions. The first proposal, which considers lexical aspect, invites exceptions mainly in the stative class—stative intransitive verbs that denote events. The second proposal deals with agentivity and most exceptions fall in the active class, particularly internally-caused bodily functions. However, the number of exceptions is still fewer than when lexical aspect was considered as the primary semantic factor. The third proposal, which combined lexical aspect and agentivity, admits the fewest exceptions. While I argued that the third proposal has the most empirical coverage since it is able to correctly classify most intransitive verbs in Crow, the second proposal is perhaps more suitable for use in language education. Regardless, the high reliability of the proposals in predicting verb class membership provides strong evidence that verbs in Crow are semantically determined.

Beyond just looking at the semantics of the two verb classes, I have also examined the distinct morphosyntactic behaviors that active and stative intransitive verbs exhibit. I have presented (morpho)syntactic evidence from Crow in support of the unaccusative hypothesis based on four sources of evidence. The diagnostics involving multiple exponence and noun incorporation show that subjects of active intransitives behave syntactically like subjects of transitive verbs, and that subjects of stative intransitives behave like objects of transitives. The diagnostics involving the inchoative and direct causative demonstrate that these morphemes can only attach to stative intransitive verbs. The analysis put forth to explain these restrictions argues that these two morphemes require that an agentive (external) argument be lacking from the verb's argument structure. Their sensitivity to the verb's argument structure thus reveals insights on the syntactic configuration of active and stative intransitives. The distinct syntactic patterns between active and stative intransitive verbs therefore provide support for the unaccusative hypothesis—in Crow, active intransitives are unergatives and stative intransitives are unaccusatives.

References

Boyle, John P. 2007. *Hidatsa morpho-syntax and clause structure*. Chicago, IL: University of Chicago dissertation.

Dowty, David. 1979. *Word meaning and Montague grammar*. Dordrecht, Netherlands: Reidel Publishing Company.

Duranti, Alessandro. 2004. Agency in language. In Alessandro Duranti (ed.), *A companion to linguistic anthropology*, Malden, MA: Blackwell.

Golston, Chris, John Boyle & Lewis Gebhardt. 2018. Crow has no incorporation. Paper presented at the Annual Linguistic Society of America Conference. Salt Lake City, UT.

Gordon, Ray & Randolph Graczyk. n.d. A dictionary of Crow. Ms.

Graczyk, Randolph. 2007. *A grammar of Crow*. Lincoln, NE: University of Nebraska Press.

Harris, Alice. 2017. *Multiple exponence*. Oxford, UK: Oxford University Press.

Jackson, Eric M. 2005. *Resultatives, derived statives, and lexical semantic structure*. Los Angeles, CA: University of California, Los Angeles dissertation.

Ko, Edwin. 2019. Active-stative agreement in crow. Paper presented at the Society for the Study of the Indigenous Languages of the Americas Summer Meeting. Davis, CA.

Lakoff, George. 1966. *Stative adjectives and verbs in English* (NSF Report no. 17). Cambridge, MA: Computational Laboratory, Harvard University.

Legendre, Géraldine & David S. Rood. 1992. On the interaction of garmmar components in lakhota: Evidence from split intransitivity. In Laura A. Buszard-Welcher, Jonathan Evans, David Peterson, Lionel Wee & William Weigel (eds.), *Proceedings of the eighteenth annual meeting of the berkeley linguistics society*, 380–394. Berkeley, CA: Berkeley Linguistics Society.

Levin, Beth & Malka Rappaport Hovav. 1995. *Unaccusativity: At the sytax-lexical semantics interface*. Cambridge, MA: MIT Press.

Li, Chao. 2007. Split ergativity and split intransitivity in Nepali. *Lingua* 117(8). 1462–1482.

Mithun, Marianne. 1991. Active/agentive case marking and its motivations. *Language* 67(3). 510–546.

Perlmutter, David M. 1978. Impersonal passives and the unaccusative hypothesis. In Jeri J. Jaeger, Anthony C. Woodbury, Farrell Ackerman, Christine Chiarello, Orin D. Gensler, John Kingston, Eve E. Sweetser, Henry Thompson & Kenneth W. Whistler (eds.), *Proceedings of the fourth annual meeting of the berkeley linguistics society*, 157–190. Berkeley, CA: Berkeley Linguistics Society.

Pustet, Regina. 2002. Split intransitivity revisited: Comparing Lakota and Osage. *International Journal of American Linguistics* 68(4). 381–427.

Rankin, Robert L. 2004. A diachronic perspective on active/stative alignment in Siouan. Ms. Lawrence, KS: University of Kansas.

Rice, Keren. 1991. Intransitives in Slave (Northern Athapaskan): Arguments for unaccusatives. *International Journal of American Linguistics* 57(1). 51–69.

Schwarz, Martha. 2019. Personal communication.

Wallace, Karen K. 1993. *Verb incorporation and agreement in Crow*. Los Angeles, CA: University of California, Los Angeles dissertation.

West, Shannon L. 2003. *Subjects and objects in Assiniboine Nakoda*. Victoria, BC: University of Victoria M.A. thesis.

Williamson, Janice S. 1979. Patient marking in Lakhota and the unaccusative hypothesis. In Paul R. Clyne, William F. Hanks & Carol L. Hofbauer (eds.), *Papers from the fifteen regional meeting, Chicago Linguistic Society*, 353–365. Chicago, IL: Chicago Linguistic Society.

About the Proceedings

The *Proceedings of the Siouan and Caddoan Languages Conference* (SCLC) is an annual periodical where any presenters at the SCLC are welcome to submit papers for publication. The scope of the papers that appear here range from highly theoretical linguistics papers to papers dealing with philology, language education, and cultural topics relating to Siouan- or Caddoan-speaking peoples. The aim of the *Proceedings* is to be inclusive towards scholars from a variety of backgrounds and provide a venue for work that scholars, community members, and/or educators can utilize to promulgate their work relating to these two language families.

These papers are reviewed and proofed by each volume's editor(s) and other volunteers from the Siouanist and Caddoanist communities. These volunteers are typically sought through the Siouan Linguistics Listserv. While the Listserv may have "Siouan" in the title, it is also a community for those working on Caddoan languages, given the historical overlap between scholars who worked with communities from both cultural and linguistic groups.

Anyone interested in Siouan and/or Caddoan languages is welcome to sign up for the Siouan Linguistics Listserv, which is operated through the University of Nebraska. The link to sign up for the Listserv appears below:

listserv.unl.edu/signup-anon?LISTNAME=siouan&LOCKTYPE=LIST&REALNAME=TRUE

Volumes of the *Proceedings* from the SCLC 38 and onward are available at http://www.siouan.org, a website dedicated to archiving information related to the SCLC, as well as compiling bibliographies of published worked on Siouan and Caddoan languages.