

PRACTICAL AND ETHICAL ISSUES IN LEXICOGRAPHY
EXAMPLES FROM THE CREEK DICTIONARY PROJECT

Jack B. Martin and Margaret McKane Mauldin

College of William & Mary and University of Oklahoma

1. Most field methods classes in graduate linguistics programs focus on transcription, organization, and analysis of data from a speaker of another language. Relatively little attention, if any, is devoted to the political, social, and ethical dimensions of fieldwork and scholarship. Ethical issues are rarely felt to be a matter of concern when fieldwork is conducted on major languages. As linguists increasingly turn their attention to endangered languages (which are endangered precisely because they are spoken in marginalized communities), there is a heightened need to find solutions to ethical and practical issues and to find solutions quickly.

This paper is a frank and personal description of some of the problems we faced in the seemingly straightforward task of writing a bilingual dictionary of Creek and English. Martin has conducted fieldwork as an outsider in communities in Oklahoma, Texas, and Florida. Mauldin has served as a Creek language consultant in a field methods class, has conducted fieldwork as an insider in her own community, teaches Creek at the University of Oklahoma, and has been involved in collaborative work preparing texts, interviewing other speakers, and in writing a dictionary. In 2 we discuss our own backgrounds and give a brief overview of the methods used to compile the Creek dictionary. In 3 we describe specific issues arising in the course of our fieldwork that we hope will be of interest to other researchers. While dictionaries are frequently reviewed in linguistic journals, it is remarkable how few of the issues we address are discussed in print.

2. Creek is a Muskogean language indigenous to Alabama and Georgia but which, through forced and partly voluntary migration, is now spoken in three communities: the Muskogee (Creek) Nation of east-central Oklahoma (with about 3,000 speakers out of a population of 40,000), the Seminole Nation of central Oklahoma (with approximately 1,200 speakers out of 11,900), and the Seminole Tribe of Florida (with a few dozen speakers at the Brighton Reservation in central Florida).

Creek has a fairly old written history compared to other languages in the family. Literacy developed in the 19th century when missionaries (most notably A. E. W. Robertson and her husband William S. Robertson) worked with native speakers to develop a practical writing system, to teach in Creek (through an interpreter), and to publish materials in Creek. Their former students subsequently gained prominence within the Muskogee Nation, producing laws, letters, court documents, and other materials in Creek.

The traditional spelling of the consonants in Creek is phonemic, differing from standard Americanist transcription in that *r* represents a voiceless lateral fricative.

TRADITIONAL	PHONEMIC.
<i>p t c k</i>	<i>p t č k</i>
<i>m n</i>	<i>m n</i>
<i>f s</i> <i>h</i>	<i>f s</i> <i>h</i>
<i>r</i>	<i>ʃ</i>
<i>w l y</i>	<i>w l y</i>

There is a weaker correspondence between traditional and phonemic spellings of vowels and diphthongs, however

TRADITIONAL	PHONEMIC
<i>e</i>	/i/
<i>ē</i>	/i:/
<i>u</i>	/o/
<i>o</i>	/o/ or /o:/
<i>v</i>	/a/
<i>a</i>	/a:/ or /a/
<i>au</i>	/aw/
<i>ue</i>	/oy/

The mismatch in the vowels arises because many speakers find it easier to represent vowel quality (tense vs lax) rather than vowel quantity (short vs long). Contrasts in nasality, tone, and stress are also not indicated in the traditional spelling.

Today most fluent speakers of Creek are over 60, only a few individuals in their teens and twenties are able to speak the language, and only a handful of individuals read and write easily. As such, Creek is considered an endangered language. As the number of fluent speakers has decreased in this century, there has been a growing interest in documenting and attempting to maintain the language. The most substantial body of linguistic research on Creek in this century was conducted by Mary R. Haas in the 1930s and 1940s, resulting in several articles, unpublished texts, and an unpublished vocabulary. Other linguists who have published articles on Creek since then include Karen Booker and Donald Hardy. During the 1970s, Susannah Factor worked with the Creek-Seminole Bilingual Education Project to develop new Creek primers. In 1991, the University of Oklahoma began offering three levels of Creek courses (taught since 1995 by Margaret Mauldin). The Seminole Nation and the Muskogee Nation established language committees in 1993 and 1996, respectively. Several public school districts in Oklahoma began teaching Creek as a second language in 1993. The Oklahoma Native American Language Development Institute received a three-year grant from 1992 to 1994 to instruct native speakers of indigenous languages in linguistics and to develop curricula. Claudette Robertson, George Bunny, and Ted Isham founded the Muskogee Language Institute in 1995, and George Bunny began teaching Creek at Oklahoma State University and Oklahoma City University in 1996. Margaret Mauldin established a group of youngsters in 1996 to study and pass on Creek hymns, and she has begun publishing collections of Creek hymns with tapes. In all, it is probably safe to say that there are probably more people actively studying Creek now than at any time in the past.

Jack Martin began working on Creek in 1986 in a field methods class on the Oklahoma Seminole dialect taught by Pamela Munro with Joanna Freeman. Most of his early years were spent learning to hear Creek tone and stress and trying to understand Creek grammar. His interests changed dramatically in 1991 when, through John Moore and Morris Foster of the University of Oklahoma, he met Margaret Mauldin in her home town of Okemah. Mauldin subsequently attended the Oklahoma Native American Language Development Institute in 1992 and became an instructor the following two years. We subsequently began collaborative work on a number of projects. In 1994, we received a grant from the National Endowment for the Humanities to create a new Creek dictionary.

The methods we have been using to write the Creek dictionary are traditional ones applied in a collaborative way: a) we selected a corpus of texts, b) we made concordances of these texts, and, c) we searched the concordances for new words, new senses, and

example sentences to include in the dictionary. We supplemented this material by checking and adding words from Loughridge and Hodge's 1890 dictionary, Haas's vocabulary and other sources, and by asking everyone we knew to help us. We were also greatly influenced by three recent dictionaries of Muskogean languages (Sylestune et al. 1993, Munro and Willmond 1994, and Kimball 1994).

Creating the corpus was by far the most time-consuming aspect of the project. While monolingual dictionaries are customarily compiled by searching texts, bilingual dictionaries of major languages are usually written by consulting monolingual dictionaries or by relying on a native speaker's intuition. Hartmann (1987) makes the quite sensible suggestion that a bilingual dictionary should be compiled by consulting a corpus of parallel texts. In this way, the compiler has access to specific choices that skilled, bilingual translators have made in matching words between languages. Pursuing this approach, we set about gathering a range of letters, stories, and other materials from the 19th century to the present. Mauldin added translations and tape-recorded herself reading each document. Martin then added phonemic transcriptions based on the tapes and entered the material in a machine-readable format using a program called Interlinear Text (IT).¹ The result is a textbase of parallel language files with separate fields for: a) the target language (in a normalized traditional spelling), b) the phonemic transcription, and, c) the translation.

For the concordance, we used a program called Conc. Conc has the ability to import files from IT and then to generate concordances based on words in specific fields (see Sample 1 at the end of this paper). In our case, we made concordances based on our normalized traditional spelling of Creek texts. Martin then went through the concordance searching for new words and senses, printing long lists of words that Mauldin then checked. Mauldin added or revised definitions for these words, which were subsequently added to the dictionary.

3. We have been so busy collecting and analyzing materials for the last three years that we have not always had time to step back and consider the implications of our work, the decisions we made along the way, or the impact our work might have on Creek-speaking communities. We discuss these issues in the following subsections.

Issues relating to orthography. By far the most complicated issue we dealt with was the issue of spelling. Linguists who work on languages without long written histories have the freedom to design their own writing systems and generally favor phonemic systems of transcription, sometimes with the additional requirement that the orthography be typable (Munro 1995). Creek has a written history, however, and even though literacy is low, people are used to the appearance of the traditional orthography. Because the New Testament is written in this system, some people view attempts to change the writing system as blasphemy. Others worry that the traditional spelling might be too hard for children to learn in schools. A further concern of ours was whether making radical changes in the spelling system would shift authority in matters relating to language from community members to linguists. After extensive discussions with many individuals (including especially Akira Yamamoto, David Skeeter, Tim Thompson, and Mekko Lewis, whom we thank), we opted to retain the traditional Creek spelling, but to add phonemic transcriptions (in Haas's orthography) for each entry word and example sentence in the dictionary. This practice has the effect of validating traditional practices, connecting the past with the present, and building on the existing abilities of community members to read.

¹Interlinear Text, Conc, and other useful software for various platforms may be downloaded from the Summer Institute of Linguistics at www.sil.org. We are grateful to the developers of these applications for their willingness to share their products.

A related problem involves word spacing. Words in Muskogean languages can be quite long. The solution in the 19th century was to add spaces after morphemes that would today be treated as prefixes. This makes words easier to read and allows nonnative speakers to more quickly identify the root and find it in a dictionary. The 20th century dictionaries of Alabama, Chickasaw, and Koasati all reject this approach, preferring to write phonological words as single units. We have taken an intermediate path, adding spaces after only a few prefixes. To make words more readable in the dictionary, we have divided entry words into syllables.

The form and structure of entries. Special problems arise in dictionary work on languages with rich morphology that are not typically discussed in the work on major languages. One issue is the form of an inflected word to select for the main entry. This issue can be especially complicated in languages with extensive prefixing (Munro 1995). In our case, however, Loughridge and Hodge (1890) and Haas had already established the practice of listing Creek verbs in a particular form. We chose not to diverge from their practice.

Another difficulty involves deciding whether to include lexemes as main entries or as subentries. To take just one example, Creek has several verbal prefixes that function much like English prepositions, as the following Creek forms show.

opvnetv /opan-itá/ to dance
 em opvnetv /im-opan-ita/ to dance for (someone)
 iem opvnetv /a -im-opan-itá/ to dance with (someone)
 'sopvnetv /i(s)-opan-ita/ to dance with (a feather, etc.)

After much discussion, we decided to list some derived forms as subentries to the base form, as in the following entry.

o pv ne tv /opan-itá/ to dance
 em o pv ne tv /im-opan-ita/ to dance for
 iem o pv ne tv /a -im-opan-itá/ to dance
 with (someone) 'so pv ne tv /i(s)-opan-
 ita/ to dance with (a feather, etc.)

This is the form Mauldin prefers. She reports she would normally think to look up the derived forms as main entries, and she likes to see relationships between words.

Munro and Willmond (1994) favor main entries over subentries, as the abbreviated Chickasaw entries below show.

hilha to dance
 jhilba to dance for
 sahilba to dance in, dance at
 baahilha to dance with

Kimball (1994) follows Munro's practice in the Koasati dictionary.

bitim to dance
 onabitim to dance around something

Sylestine et al (1993) go further in listing the same information twice in their Alabama dictionary, once in a subentry and once in a main entry.

bitli to dance
 ibaabitli to dance with
 ibaabitli to dance with bitli

This might not be an option in a larger work.

We saw advantages and disadvantages in each of these solutions when there are irregularities in the morphology, adopting an approach favoring main entries allows greater systematicity in the dictionary. The Creek verb *em pvl̄sēcetv* /im-palsi c-utá/ 'to mate' never occurs without *em* /im-/, for example, so we listed it under *em*, with a crossreference

em pvl̄sēc̄ ce·tv /im-palsi c-utá/ to mate
 -pvl̄sēc̄·ce·tv see *em pvl̄sēcetv*

This makes sense to a native speaker (who would never consider looking the word up under the second form). The nonnative speaker unfortunately must occasionally be referred to other entries, but in so doing learns which words have obligatory prefixes. In short, decisions regarding the structure of entries were difficult, but aside from orthography, no other decision other than orthography has such a profound effect on users.

Making words easier to find. A few dictionaries (such as *A Dictionary of Everyday Crow*) are organized by semantic field (Animals, Body Parts, etc.). This practice makes it easier for speakers who do not read and write well to find words in the dictionary, and may help teachers of the language in developing lesson plans. It also allows native speakers to see a list and reflect on what might be missing.

We opted instead for a traditional alphabetical organization for the Creek dictionary. This is partly because the previous Creek dictionary (Loughridge and Hodge 1890) had this arrangement, and because many speakers are at least familiar enough with the traditional spelling to be able to find words. In Florida (where the rate of literacy is lower) we have been producing trilingual words lists (English-Creek-Mikasuki) arranged by semantic field.

A related issue concerns the reverse dictionary or index. A fairly common practice (made more popular by the existence of lexical database programs) has been to concentrate on writing, e.g., a Creek-to-English dictionary and including only an abbreviated index for the English-to-Creek section. While limitations of time and money may force us to adopt this solution, we noticed that many people find it easier to find words by the English definition (since they better control English spelling). We are still writing the English-to-Creek section of the dictionary, but we would like it to include much of the same information found in the Creek-to-English section.

Inclusion of geographical variants. When we wrote our original proposal, we planned to concentrate on forms of Creek spoken in Oklahoma. Reviewers suggested that we extend coverage to all three Creek-speaking communities. There were political ramifications to consider in this regard. Muskogee speakers frankly have a tendency to look down upon the speech of Seminoles, much the way some speakers of standard American English devalue Appalachian varieties of English. By including Seminole forms, we worried that we would cause Muskogee speakers to question the accuracy of the work. Conversely, by excluding Seminole variants, we worried that we might be accused of contributing to the political fragmentation of the three speech communities. In the end, we decided to base the dictionary on Muskogee and to use only Muskogee example sentences, but to include notes on Seminole forms, as in the following entry.

cu fun wv /cofónwa/ 1 a pointed tool, 2.
Musk needle, awl, 3. *Okla* fork, 4. *Fl*
 pitchfork [cf *Sem* *esrapoitv* /is-lopó tt-a/
 'needle', *Fl* *cufun-yakpe* /cofon-yákp-*l*
 'fork']

Seminole may still feel underrepresented, but we have tried to provide symbolically for their inclusion in the project.

Language and dialect names One problem we faced was that there was no agreement within the three communities on the name of the language or the names of dialects. Muskogees say they speak Creek or Muskogee (also spelled Muscogee, Mvskoke, etc.) Oklahoma Seminoles say they speak Seminole. Florida Seminoles at Brighton say they speak Creek, associating the term *Muskogee* with Oklahoma. The term *Seminole* for Florida Seminoles could refer to either of the two languages spoken within the tribe. We thus found it necessary to develop a consistent terminology that could be applied to all three communities. Since *Creek* is felt by some to be a broader term than *Muskogee*, we opted to refer to the language all three groups share as *Creek*. We anticipate that there will be some individuals (perhaps especially Oklahoma Seminoles) who dislike this terminology, but we were unable to please everyone.

The issue of privacy Issues related to privacy arose in compiling the Creek corpus and dictionary. Since language is inevitably shaped by everyday use, we wanted to include conversational material. Since many conversations are of a highly personal nature, and some may even include names and criticisms of other community members, we decided not to include them in our corpus. We would like to return to this topic once our work is better understood in the community.

In our dictionary, as in the Alabama and Chickasaw dictionaries (Sylestine et al 1993, Munro and Willmond 1994), we included the initials of speakers who had idiosyncratic ways of speaking or who knew words that could not be verified from others. This is exemplified in the following entry from the Creek dictionary.

co·kv-tv·lv me·ci cv, co kv-tv lv mē cv
 (B), co kv-tv le·mi cv (LM) /co-ka-
 talamicéyc-a, -mu c-a, -liméyc-a/ newspaper
 editor

Dictionaries of major languages do not include an individual's words in this way, but many of us work in speech communities that have contracted. It is quite possible that an individual's words preserve a variant that was once in wider use. Where individuals understand publishing and dictionaries, it is easy to ask permission to include the speaker's initials, but what if individuals do not understand the project well, or have not thought about how other people would react to their work with outsiders? What if the individual dies before permission is obtained? We sought to avoid the temptation to include every single variant in our dictionary. We know that researchers in the future will have access to our fieldnotes, and so felt comfortable leaving out variants where there were doubts about privacy or accuracy.

Another issue relating to privacy involves the inclusion of unique personal names or nicknames in the dictionary when these are not historical figures. Kimball (1994) chooses to include these, but we did not feel it was appropriate.

The issue of ownership Most of the people who helped us were happy to have their stories included in the corpus. We were uncertain how to draw on the extensive work of the

late Mary R. Haas, however. Martin's understanding from a conversation in 1993 was that Haas felt uncomfortable publishing the texts she had gathered—all of the Creek texts need to be checked, she said, but checking would now be difficult because modern speakers no longer speak the same way her consultants did in the 1930s and 1940s. While we were therefore reluctant to use her materials, we also knew that linguists in the future would not hesitate to do so—a distance in time and place between researchers makes it easier to treat a work philologically. After much consideration, we decided we would search Haas's texts to locate new words for checking, but we decided not to include any example sentences from Haas's materials. We believe Haas would have found this solution acceptable, but we worry that we may have underrepresented the value of her materials in the process.

Another issue related to ownership concerns publication of the dictionary (and copyright). Because academics are often evaluated by the prestige of publishers, there is little incentive to find local publishers that might be of greater benefit to the group. In our case, the Muskogee Nation did not have its own publishing facilities, so we did not consider publishing through the tribe. An intermediate solution might be to publish drafts of the dictionary locally and the final edition through a larger press.

The inclusion of religious or ceremonial information. There is a tradition in dictionaries of American languages of including information on medicinal uses for plants and cultural information that would be of interest to those outside the community. Some people in Creek-speaking communities feel strongly that words relating to medicine should not pass outside of the community, however. When Howard and Lena's *Oklahoma Seminoles: Medicines, Magic, and Religion* appeared, Willie Lena was widely criticized for teaching medicine to an outsider. The perception was that the traditions of the community had been sold for Howard's gain. As a result, we decided not to investigate this area thoroughly and recorded words relating to medicine in our notebooks only when they came up naturally. We also decided not to include sensitive information of this kind in the dictionary or in example sentences.

Descriptive vs. prescriptive lexicography. Linguists routinely teach their students that all varieties of speech are legitimate. The usual assumption among nonlinguists, however, is that a dictionary is a guide to correct (even hypercorrect) speech. These different assumptions initially led to disagreements between us. Martin tended to write words in contracted ways (the way words were pronounced in everyday conversation), while Mauldin preferred spellings that clarified the origin of words. In other instances, Mauldin felt Martin's phonemic transcriptions were too abstract. Agreement was reached by aiming for a higher register, and by allowing small differences between the traditional orthography and the phonemic representations. We also deleted forms that, although used, were felt to be childish pronunciations or simply wrong. These forms will, of course, be available in our notes, but will not be in the published work where they might weaken the authority of the dictionary.

Representing different genres. We have noticed a tendency for linguists and anthropologists to favor the collection of legends and myths when they work on American languages. We felt this practice might have a distorting influence on descriptions of Creek. First, works of this kind may be heavily stylized. Second, many stories are designed for children and tend to simplify aspects of the language. For this reason, we tried in our corpus to include a large number of letters and other materials for balance.

We have also noticed a tendency to romanticize indigenous traditions at the expense of modern life. Example sentences derived from salvage work and rooted in assumptions about old customs perpetuate stereotypes. By creating examples of modern life, we hoped that the reader would be made aware that Creek is a living language.

Accuracy vs size The previous dictionary of Creek (Loughridge and Hodge 1890) is valuable, but filled with errors. We know that our dictionary is not error-free, but we would rather have a smaller, more accurate dictionary than a larger, less accurate work. We have noticed that inaccurate works appearing in languages with few written materials lead community members to question the value of written materials.

Choosing language authorities An important issue affecting fieldwork is the problem of matching suitable native speakers with suitable linguists and/or training native speakers to be linguists. When Martin began working on Creek, he naturally had little experience with the language. It would have been difficult or arrogant for a graduate student in his position to approach one of the tribal organizations to propose collaborative projects. Instead he asked for permission to conduct his research, and then proceeded to locate individuals in the community who seemed to him to be good speakers. After several years of research, he felt more comfortable about his knowledge of the language and began meeting more people in the community (including Mauldin).

As one of the younger Creek speakers, Mauldin had similar concerns. In a society in which age is often used to measure fluency and authenticity, Mauldin felt a need to establish her authority through education (by taking linguistics courses and studying Creek as an object of grammatical analysis).

By working on the language for a number of years, studying it, and producing materials, we were slowly able to raise our status in the community. We then found it easier to approach tribal organizations to discuss language-related projects. Tribal organizations are now more willing to take us seriously and have responded by establishing panels of language experts to review our work. In retrospect, though, our practice might have been disastrous, largely because the tribal organizations involved never had the opportunity to accept or reject us. Established scholars have an easier time offering their help to organizations—it is not clear to us how younger scholars should proceed.

Relations with tribal governments We conducted our research largely independent of tribal governments, though we have promised to submit our work to language committees for their suggestions and approval.

Profit One sensitive topic is the matter of profit from work of this kind. Profits arise from the sale of a dictionary, from grants used to write it, and from the increase in status associated with publishing research. There is generally little profit in the sale of dictionary, especially considering the long period of time needed to write one. The sharing of indirect costs associated with grants is a more difficult issue: academics at universities are given a great deal of credit for writing grants through their home institutions, so that there is little incentive to writing grants through or in collaboration with tribal organizations. We hope that this sort of collaborative work will be easier in the future.

4. Conclusion.

We have tried to give a frank appraisal of our collaborative work on a new Creek dictionary. We hope that some of the issues addressed will lead to deeper reflections and better solutions in this kind of research.

REFERENCES

- Hartmann, R R K 1987 "Contrastive textology and bilingual lexicography" In James Monaghan, ed , *Grammar in the Construction of Texts*, 114-122 London Frances Pinter
- Howard, James H , and Wille Lena. 1984 *Oklahoma Seminoles Medicines, Magic, and Religion* Norman University of Oklahoma.
- Kimball, Geoffrey D 1994 *Koasati Dictionary* Lincoln University of Nebraska Press
- Loughridge, R M and David M Hodge 1890 *English and Muskokee Dictionary* Creek Mission, Indian Territory *Dictionary of the Muskokee or Creek Language in Creek and English* Red Fork, Indian Territory Reprinted as one volume by B Frank Belvin, 1964
- Medicine Horse, Mary Helen [Ishtaléshchia Báachua Héeleetaalawe] 1987 *A Dictionary of Everyday Crow* Crow Agency Bilingual Materials Development Center
- Munro, Pamela 1995 "Making a Zapotec dictionary Some general issues" Ms, UCLA
- Munro, Pamela and Catherine Willmond 1994 *Chickasaw An Analytical Dictionary* Norman University of Oklahoma Press
- Sylestune, Cora, Heather K. Hardy, and Timothy Montler 1993 *Dictionary of the Alabama Language* Austin University of Texas Press

SAMPLE 1 Example of a concordance made by exporting interlinear files from IT to Conc Here, *ohhatalá ka t* has been selected in the concordance (the second window), and the first window scrolls to the indicated passage from the text.

File Edit Font Options Layout Build Windows

Text

r| 1878-Postoak 12/11 74
 c Ohhvrvlakat cawvnrw cem apohucaket
 p **ohhatalá ka t** ca wánwa cím= a.poheycá kit
 e Further, I have heard that my sister is being obedient in your guidance

r| 1878-Postoak 12/11 75
 c momet ecēfvcevcicaket arvcokeñ pohyat vt afvckēt omes
 p mo mít icl faciceycá.kit a.lacókm po.hayá.t.ac= a.fácku.t.ó.mis
 e and that makes me very happy

1874-1905 TU conc

1878-Postoak	ómho yáñks cí. [?]	ohhatalá ka t má ístu aksomiceyhó canks ma
1878-Postoak	pona.yá.c.kit omáñ s cí.	ohhatalá ka t ca wánwa im= tá'k'koséys aho
1878-Postoak	ko mayít ó mis	ohhatalá ka t ca wánwa cím= a.pohewá.kit
1878-Postoak	láykm oponaká ha.yéys	ohhatalá ka t ca wánwa hí'án im= aha yíci
1878-Postoak	is= in= ponáyuckah s cí.	ohhatalá ka t ca wánwa im= akiBéycka m=
1886-Smith 2f	kó meys cí. mo mín	ohhatalá ka t istu+ma.skó ki im= ponáka íst
1886-Smith 3f	cí. yó mosn fikhonnéys	ohhatalá ka t cím= có.kan hí.cayá.t.c'vni =te
1878-Postoak	wéhkickas hopoñanka	ohhatalá ka t kím sím= pona.yícka t omáñ s cí e
1884-Tanyan .	cókpí+hámki [?]	ohhatalá ka t an= híckah s kó meys mo mín
1883-Smith 4	ná.ki ístó.mosa t am=	ohhatalá yín sí y= ato tkéys is= hóyat o me
1899-Tanyan	monkó to miká iy=	ohhatalá cím= co yéys mo mín caccóstu El