

Chahta Anumpa: A Multimodal Corpus of the Choctaw Language

Jacqueline Brixey, Eli Pincus, Ron Artstein

USC Institute for Creative Technologies
12015 E Waterfront Dr, Los Angeles, CA 90094
{brixey, pincus, artstein}@ict.usc.edu

Abstract

This paper presents a general use corpus for the Native American indigenous language Choctaw. The corpus contains audio, video, and text resources, with many texts also translated in English. The Oklahoma Choctaw and the Mississippi Choctaw variants of the language are represented in the corpus. The data set provides documentation support for the threatened language, and allows researchers and language teachers access to a diverse collection of resources.

Keywords: endangered languages, indigenous language, multimodal, Choctaw, Native American languages

1. Introduction

This paper introduces a general use corpus for Choctaw, an American indigenous language. The Choctaw language is spoken by the Choctaw tribe, who originally inhabited the southeastern United States. The tribe is the fourth largest indigenous group by population in the United States with 220,000 enrolled members.¹ The Choctaw language, however, is classified as “Threatened” by Ethnologue,² as there are only 10,400 fluent speakers and the language is losing users.

While sporadic written records of the language appear as early as 1715 (Sturtevant, 2005, page 16), systematic writing of Choctaw only began with the arrival of American missionaries, led by Rev. Cyrus Byington in 1819 (introduction to Byington, 1870). The first text published in Choctaw was the Bible, and this is still the longest published text in the language today. Efforts during the 1900s aimed to forcefully assimilate Native Americans and suppress indigenous languages (Battiste and Henderson, 2000). For this reason, many Native Americans did not learn their ancestral language, and few works are publicly published in these languages. As a result of this history, the Choctaw language has few published works and little text representation online.

While work has been undertaken to document the language and conduct linguistic studies, no digital corpus of the Choctaw language exists. The present work has gathered sparse text resources representing different variants of the language from teaching materials, books, and scholarly material. Video and transcribed audio clips where the Choctaw language is spoken in complete phrases are also included in this corpus. The majority of data in this work has an accompanying English translation, with a smaller portion in monolingual Choctaw.

For most of the world’s languages, no data in a machine-readable format are available for human language technology applications. Compared to well-resourced languages such as English and French, natural language processing tools for Native American indigenous languages are still

not abundant. The goal of this database is thus to first preserve a threatened language. The second contribution is to compile a comprehensive data set of existing resources for novel research opportunities in history, linguistics, and natural language processing. The final contribution is to provide documentation of the language for language learners and teachers, in order to assist in revitalization efforts.

2. Choctaw tribe and language

2.1. People and dialects

The Choctaw tribe originally resided in the southeast of the United States, in what today would be Alabama, Louisiana, and Mississippi. In the early 1830s the Choctaws were forcibly relocated to Oklahoma in the migration known as the “Trail of Tears”, though some remained in Mississippi. Today there are three Federally recognized Choctaw tribes:³ Jena Band of Choctaw Indians (in Louisiana), Mississippi Band of Choctaw Indians, and The Choctaw Nation of Oklahoma. Speakers are concentrated primarily in Mississippi and Oklahoma (Ulrich, 1993), but also in Alabama, Louisiana, Texas, and California. For many speakers in Oklahoma (Williams, 1999), Choctaw is their second language, and revitalization efforts have worked to establish language courses at local schools. Choctaw is spoken by all ages in Mississippi, but is losing speakers over time. The language belongs to the Western Muskogean language family; it is most closely related to Chickasaw, and also related to the Creek and Seminole languages (Haas, 1979). Dialectal variation in Choctaw has been the matter of some debate (Nicklas, 1972; Broadwell, 2005; Broadwell, 2006): while sources agree that there are three dialect variants in Mississippi which reflect original settlement patterns, it is unclear whether and to what extent similar variation has been carried over to Oklahoma. Broadwell (2006) identifies four present-day regional variants: Mississippi Choctaw, Oklahoma Choctaw, Louisiana Choctaw, and Mississippi Choctaw of Oklahoma; the latter is spoken by Choctaws who live in The Chickasaw Nation in Oklahoma, and are believed to have been relocated there from Mississippi in

¹<https://web.archive.org/web/20120512040555/http://www.ok.gov/ojac/documents/2011.FINAL.WEB.pdf>

²<https://www.ethnologue.com/language/cho>

³<https://www.federalregister.gov/documents/2017/01/17/2017-00912/indian-entities-recognized-and-eligible-to-receive-services-from-the-united-states-bureau-of-indian>

| Consonants | | | | | | | | | | | |
|------------|---------------|---|------|-----------------|-------|------|------------------|---|-----|-----------|-------|
| p | b | t | k | f | s | h | m | n | l | w | y [j] |
| [tʃ] | ch, č | | | [f] | sh, š | | | | [ɬ] | hl, lh, ł | |
| Vowels | | | | | | | | | | | |
| [a] | a, v, v | | [i] | i | | [o] | o, u | | | | |
| [a:] | a, á, aa | | [i:] | e, í, í, ii, ie | | [o:] | o, ó, oo | | | | |
| [ã] | ã, an, am, ạ | | [ĩ] | ĩ, in, im, ị | | [õ] | õ, u, on, om, ọ | | | | |

Figure 1: Choctaw sounds and orthographic variants.

the early 1900s. Overall, regional variation in Choctaw is fairly minor, with some variation in phonetic detail (Ulrich, 1993) and a small number of lexical differences (for example, the word for “onion” is typically *hatofulaha* in Oklahoma and *shatshonna* in Mississippi). A much larger source of variation in our corpus is differing orthographic conventions.

2.2. Sounds and orthography

The Choctaw language has 15 consonants, and 9 vowels in three series: short, long, and nasalized; the orthography uses the Latin script, but is not fully standardized (Figure 1). Broadwell (2006, section 1.2) identifies three broad writing systems: traditional, Mississippi, and modified traditional; our collected materials, however, show more variation. Generally, Oklahoma speakers write the short and long vowels as *v-a*, *i-e*, and *u-o* (with *v* often substituting for *u*), while Mississippi speakers use *a-á*, *i-e*, and *o-ó*. Mississippi speakers consistently write the lateral fricative [ɬ] as *lh*, while Oklahoma speakers use *lh* before a consonant and *hl* otherwise. In both variants, long vowels are sometimes but not always doubled, and nasalized vowels are sometimes represented with a macron below and sometimes with a following nasal consonant character. Broadwell (2006) characterizes the use of *č*, *š*, *ł*, *ạ*, *ị*, and *ọ* as Mississippi Choctaw, but we have found it primarily in instructional materials rather than general Mississippi use.

In addition to the above conventions which are in wide use, our sources include idiosyncratic variants that are limited to a single work: The dictionary of Byington (1915), published nearly 50 years after his death, has unique symbols introduced by the editor: *ạ* for short [a], and a superscript for nasalized vowels *aⁿ*, *iⁿ*, *oⁿ*. Nicklas (1972) represents long vowels with a macron above *ã*, *ĩ*, *õ*.

Choctaw has additional sounds which are not indicated orthographically: a glottal stop whose phonemic status is unclear (Broadwell, 2006, section 2.6), and a lexically specified pitch accent in certain words and inflections (Nicklas, 1972, section 1.4). The glottal stop and pitch accent are rarely if ever contrastive and are not indicated in ordinary writing, but some scholarly or instructional materials in the corpus indicate them by an apostrophe or accent mark.

In our corpus, each entry retains the orthographic convention of its source.

2.3. Morphology and syntax

Choctaw word order is subject-object-verb, and adjectives follow the nouns they modify. Choctaw also has a com-

plex morphology system, including infixation. The following examples illustrate some of the morphology features of Choctaw.

1. *Ikana ut aiittutoba ia tuk*
Ī-kana ut aiittutoba ia tuk
his-friend the store go PAST
His friend went to the store.
2. *Ashekonopa ilupput tahakchi li*
Ashekonopa ilupput ta<ha>kchi li
knot this tie<quickly> 1SG
I tie this knot quickly.
3. *Chik impo*
Chik imp-o
not.2SG eat-NEG
You are not eating.

Orthographic conventions vary regarding the segmentation of words and morphemes. Early works tended to segment words into shorter units than is customary now (Broadwell, 2006), but even today segmentation is not consistent: some texts show agglutination of all morphemes (Nicklas, 1979), while others specifically call for separation of some morphemes (The Choctaw Nation of Oklahoma Dictionary Committee, 2016).

2.4. Reference works

The language has three reference grammars, the first by the missionary Cyrus Byington (1870, published shortly after his death), the second by Thurston Dale Nicklas (1979), and the final by George Aaron Broadwell (2006). Several dictionaries also exist, the first written by Cyrus Byington (1915), while a more recent dictionary was released by The Choctaw Nation of Oklahoma (2016).

3. Data Collection

The long-term goal of this project is to continually add newly published and newly discovered data in the Choctaw language to this corpus, to encourage language research and preservation. All work was approved for study by the Choctaw Nation of Oklahoma, with whom the first author is an enrolled member.

The current repository was formed using printed written and oral teaching materials, examples from academic research sources, audio clips, and videos. In order to adhere to copyright standards, all text entries include the appropriate identifying source information, such as author and publication title. We are currently in the process of seeking permissions to make this corpus publicly available.

All text data was manually entered into a Microsoft Access database (Figure 2). Published teaching material was gathered from the Choctaw Nation of Oklahoma and the Mississippi Band of Choctaw Indians. Marcia Haag and Henry Willis’s two books of teaching material, poetry, short stories, and correspondence formed a large portion of the Oklahoma Choctaw portion of the corpus (Haag and Willis, 2001; Haag and Willis, 2007). Archives from the Los Angeles Unified School District Indian Education Program provided teaching materials and scholarly resources for the Oklahoma and Mississippi variants.

| | Choctaw | English | dialect | Originating resource |
|----|--|---|---------|----------------------|
| 2 | Ishishkoyo? | Do you drink? | OK | Ulrich (1993) 433 |
| 3 | Oka ilappat lashpa kiyo, kapassa. | This water is not hot, it's cold. | OK | Nicklas (1971) 7 |
| 4 | Oka ilappat lashpa kiyo, pi libisha. | This water is not hot, it's just warm. | OK | Nicklas (1971) 7 |
| 5 | Tamaha iya hokak osh, nanna kiya ikchompo ki-tuok. | He went to town, but he didn't buy anything. | OK | Nicklas (1971) 8 |
| 6 | Tamaha iya hokak on, anttali-tok. | He went to town, but I stayed home. | OK | Nicklas (1971) 8 |
| 7 | Yayali-achin amahoba-tok ak osh, akyayo ki-tok. | I thought I was going to cry, but I didn't. | OK | Nicklas (1971) 8 |
| 8 | Omba-achin amahoba-tok ak on, ikombo ki-tok. | I thought it was going to rain, but it didn't. | OK | Nicklas (1971) 8 |
| 9 | Tamaha iya-tok. Yohmi hokak osh, nanna kiya ikcho-po ki-tok. | He went to town. But he didn't buy anything. | OK | Nicklas (1971) 8 |
| 10 | Tamaha iya-tok. Yohmi hokak on, anttali-tok. | He went to town. But I stayed home. | OK | Nicklas (1971) 8 |
| 11 | Yayali-achin amahoba-tok. Yohmi-tok ak osh, akyayo ki-tok. | I thought I was going to cry. But I didn't. | OK | Nicklas (1971) 8 |
| 12 | Omba-achin amahoba-tok. Yohmi-tok ak on, ikombo ki-tok. | I thought it was going to rain. But it didn't. | OK | Nicklas (1971) 8 |
| 13 | Tamaha iya-tok ak inliih kiya, nanna kiya ikchompo ki-tok. | Although he went to town, he didn't buy anything. | OK | Nicklas (1971) 9 |

Figure 2: Screenshot of how bilingual text entries are stored in the database. All entries cite the originating resource.

Published linguistic articles and grey literature – unpublished theses and manuscripts – form the section of the corpus called “scholarly sources”. In total, content in the target language was pulled from sixteen published articles, one unpublished manuscript, and two reference grammar books.

Scholarly sources and religious texts, such as hymns, were the only texts gathered from the internet. To the best of our knowledge, no blogs or web pages in only the target language exist. English texts about the history and culture of the tribe were not included in the corpus, as the primary focus of this corpus is to provide resources of the target language.

Over 580 audio files created by the Choctaw Nation of Oklahoma’s School of Choctaw Language distance learning program were compiled for this corpus. All audio clips are for learning the language, and are approximately thirty seconds long. The clips cover a variety of topics. These clips are accompanied by a transcript written by the School of Choctaw Language in both Choctaw and English. Also within the audio portion of the corpus is one five-minute-long religious clip from Global Recordings Network discovered without transcription. Finally, 13 audio files of interviews and oral traditions collected by William D. Davies were accessed from the American Philosophical Society (APS).

This corpus also includes 30 mp4 video files downloaded from YouTube that contain Choctaw speech. Video files were downloaded from YouTube using an open source downloading tool.⁴ Terms such as “choctaw”, “chahta anumpa” (the native name for the language), as well as compound terms such as “choctaw songs” were used in order to search for YouTube videos containing the Choctaw language. Each video was manually checked in order to ensure at least some Choctaw speech was present.

3.1. Future linguistic collection

As this is an ongoing effort, this initial corpus will continue to grow as more material is collected. There are numerous recordings that cannot be accessed online as well as undigitized documents in historical archives in Oklahoma and Mississippi. Due to the nature of these delicate, old, and irreplaceable items, in-person visits are required in order to access and view them. Examples of archives are those

⁴MacX YouTube Downloader

| Type | OK | MS |
|---------------------------------|-------|-------|
| Short stories | 5954 | 1693 |
| Phrases | 10911 | 331 |
| Poetry | 243 | 0 |
| Correspondence | 159 | 0 |
| Religious texts | 222 | 30010 |
| Examples from scholarly sources | 589 | 12 |
| Total | 18078 | 32046 |

Table 1: Word token counts for bilingual types of texts in two variants

held by the Choctaw Nation of Oklahoma (Durant, OK), the Sam Noble Museum (Norman, OK), and the Mississippi Band of Choctaw Indians Tribal Archives (Choctaw, MS). As funding and permissions are attained, it is expected that these remote sources will be digitized and added to this corpus.

4. Data Set

The data set comprises text, audio, and video. Texts in the data set are stored in plain text in a database. All data points list the source reference of the text with each entry in the database. Bilingual text data is separated by variant and type. The two orthographic variants – Oklahoma Choctaw (OK) and Mississippi Choctaw (MS) – are kept separate in order to make future processing easier. No Louisiana texts were found for this work. The types of text collected include short stories, single sentence phrases, and poetry (Table 1). Monolingual Choctaw data is stored separately.

4.1. Short stories

Stories and poems were all gathered from teaching materials: 29 stories from Oklahoma and 11 from Mississippi. The stories cover a range of topics, from cultural and oral traditions, to personal anecdotes. There is overlap between the MS and OK subsections. For example the story “How the Possum got its Tail” appears in both, however the story is not told identically. The MS section has 1693 word tokens, while the OK section has 5954 word tokens. The average length of an OK story is 249 words, with a minimum word count of 53, a max count of 1277 words. The average length of a MS story is 153, with a max value of 241 words, and a minimum count of 94.

| Word (English translation) | Number of tokens |
|----------------------------------|------------------|
| <i>OK</i> | |
| tuk (recent past marker) | 252 |
| mut (the) | 126 |
| ch _i (future marker) | 107 |
| ut (the) | 103 |
| li (I) | 86 |
| h _o (question marker) | 86 |
| yut (that) | 80 |
| ish (you) | 78 |
| hosh (subject marker) | 58 |
| hattak (man) | 52 |
| <i>MS</i> | |
| oklah (people) | 1049 |
| at (the) | 878 |
| alhiha (to act truly) | 720 |
| im (him/her, before vowel) | 691 |
| yat (the, who) | 643 |
| hicha (and) | 615 |
| Jesus | 545 |
| i (him/her, before consonant) | 457 |
| áchittók (he/she said) | 410 |
| ma (the, which) | 399 |

Table 2: Top 10 most frequent words in the OK and MS phrase sections

4.2. Phrases

To form the phrases portion of the data set, only grammatically complete phrases were included in the repository. A phrase is one complete sentence, such as “Iti isht bushli mut haksobali” (That chainsaw is loud), and “Impah” (She is eating). Over 2700 Oklahoma Choctaw phrases and 69 Mississippi phrases form this section. The average number of words in a phrase for both the MS and OK sections is four, with a maximum length of 15. One word representing a complete phrase in the Mississippi Choctaw of Oklahoma was found and is stored in this section of the database.

4.3. Poetry

One teaching resource (Haag and Willis, 2007) published three poems. Poetry provides novel insights into both the culture and symbology of a language, thus poetry was included in the corpus. The poetry section contains 243 word tokens from three poems.

4.4. Correspondence

Only one teaching resource (Haag and Willis, 2001) published a translated letter, with a total of 159 tokens. We expect to add to this section over time, as many letters are available on paper only and require an in-person visit to view the collection.

4.5. Religious texts

Many Choctaws today are affiliated with the Methodist, Presbyterian, or Baptist faiths. Seven OK and one MS bilingual hymns and doxologies form this section of the corpus,

with a total number of 234 word tokens. Part of Genesis⁵ and all of the Gospel of Mark⁶ were digitized by YouVersion and are in the MS orthography. The texts from YouVersion comprise 29998 tokens. Adding the remaining portions of the Old and New Testament of the Bible is ongoing.

4.6. Scholarly sources

Examples from linguistic resources are unique as they are the only input for the corpus that contain annotation for part of speech. This section contains 589 word tokens in Choctaw, and the average input is 2 words. When annotation is available, it is also included within an entry in the database, thus each entry can potentially have a Choctaw, Choctaw annotation, and English version, along with corresponding reference information.

4.7. Monolingual texts

A section of monolingual published Choctaw texts is also included in the corpus. These monolingual texts are either short stories or single phrases. This section serves as a potential test bed for future research, such as in translation or morphology detection. This section contains 1386 total word tokens, with 41 words from Mississippi, one word from Mississippi Choctaw of Oklahoma, and the remainder from Oklahoma.

4.8. Audio

All audio clips, with the exception of the religious audio retrieved from Global Recordings Network, are accompanied by a text document with the Choctaw transcription and English translation document. On average, audio clips from the Choctaw Nation of Oklahoma School of Choctaw Language are thirty seconds long, thus there are 193 minutes of audio from this resource. In each clip, the single, grammatically complete phrase is repeated twice.

Audio originally recorded by Davies for the American Philosophical Society range in length from one minute to ten minutes, for a total of 58 minutes. These recordings do have some overlap in content with the short stories section: for example, the traditional story “How the turtle got the pattern on its back” occurs in both. All recordings are accompanied by transcripts in English and Choctaw.

The final audio recording, from Global Recordings Network,⁷ is 5:40 minutes long. No transcript is provided, and no information about the speaker or recording location was given by Global Recordings Network.

4.9. Video

The 30 videos were manually annotated with one of 4 categories: Songs – 12 videos primarily containing a Choctaw song; Cultural – 3 videos that describe aspects of Choctaw culture; Story – 1 video that relates a story in Choctaw from the Choctaw culture; and Instructional – 14 videos that attempt to teach Choctaw to English speakers. The average video length is 3:51 minutes, shortest video 1:04, and the longest video is 12:29. In total, there is 105 minutes and 38 seconds of video in the current corpus. Among the videos,

⁵<https://www.bible.com/bible/1927/GEN.37.CHTW>

⁶<https://www.bible.com/bible/1927/MRK.1.CHTW>

⁷<https://globalrecordings.net/it/language/318>

16 contain bilingual content (Choctaw and English speech) while the remaining 14 contain monolingual content (only Choctaw speech).

5. Use Cases

The proposed use cases for the data set are numerous given the variety in both content and language variant. The primary use cases are for academic research in linguistics, history, and natural language processing (NLP), as well as serving to preserve the language.

Our future work is to annotate a training set of the complex morphemes of the language for study and to build a supervised model for morpheme segmentation. As the data set includes both audio and their transcriptions, an automatic speech recognition system could be developed.

One potential NLP use case is machine translation (MT). The majority of the data in our corpus is translated in English, creating a well-formed parallel data set. The language presents interesting challenges in this domain, as morphologically rich languages pose problems for MT systems from errors in word-alignment and multiple affixes. Current alignment models at word-level do not distinguish words and morphemes, and produce low-quality end translation due to misalignment (Li et al., 2016). The small size of the data set would encourage novel approaches for a MT model, as there is not enough data to use many machine learning techniques. However, as no system yet exists, a MT system would assist in generating new texts in Choctaw from English.

Cultural immersion is another use case for the corpus. Storytelling is an important means for sharing cultural norms and beliefs. We built an interactive bilingual text-based conversational agent that shares stories and parables about animals from the corpus. The agent was built using NPCEditor, a response classifier and dialogue management system (Leuski and Traum, 2011). NPCEditor employs a statistical classifier that is trained on linked questions and responses: for our agent, the questions are requests for particular stories, and the responses are the stories themselves. For example, a user might ask *Do you know a story about foxes?* and the classifier will find the most appropriate response. The agent can also make statements about itself, as well as greetings and closings to maintain dialogue flow. Some previous applications using NPCEditor include virtual museum guides (Swartout et al., 2010), a system for conversation with Holocaust survivors (Traum et al., 2015), and a Facebook Messenger chatbot to answer sexual health questions (Brixey et al., 2017)

Finally, the corpus serves as a repository for teaching and learning the language. As nearly all of the text entries are bilingual, learners and teachers alike can benefit from the translations.

6. Discussion and Conclusions

This paper introduced a multimodal data set of the low resource American indigenous language Choctaw. This data set comprises more than 50,000 word tokens in text, many with English translation, and 400 minutes of oral examples from audio and video. Future work aims to continuously

develop the data set as new publications are released. Future work will also include unpublished resources for better representation of the Mississippi Choctaws of Oklahoma, to represent the Louisiana Choctaw, and add the Bible for all available variants to the database.

7. Acknowledgements

The first author was supported by a National GEM Consortium fellowship. This work was supported in part by the U.S. Army; statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

8. Bibliographical References

- Battiste, M. and Henderson, J. S. Y. (2000). *Protecting Indigenous Knowledge and Heritage: A Global Challenge*. UBC Press.
- Brixey, J., Hoegen, R., Lan, W., Rusow, J., Singla, K., Yin, X., Artstein, R., and Leuski, A. (2017). SHIH-bot: A Facebook chatbot for sexual health information on HIV/AIDS. In *Proceedings of the SIGDIAL 2017 Conference*, pages 370–373, Saarbrücken, Germany, August.
- Broadwell, G. A. (2005). Choctaw. In Heather K. Hardy et al., editors, *Native Languages of the Southeastern United States*, pages 157–199. U of Nebraska Press.
- Broadwell, G. A. (2006). *A Choctaw Reference Grammar*. U of Nebraska Press.
- Byington, C. (1870). Grammar of the Choctaw language. *Proceedings of the American Philosophical Society*, 11:317–367. Edited by Daniel Garrison Brinton. Also published as a monograph by McCalla and Stavely, Philadelphia, 1870.
- Byington, C. (1915). *A Dictionary of the Choctaw Language*. US Government Printing Office. Edited by John R. Swanton and Henry S. Halbert. Smithsonian Institution Bureau of American Ethnology Bulletin 46.
- Haag, M. and Willis, H. (2001). *Choctaw Language and Culture: Chahta Anumpa*, volume 1. University of Oklahoma Press.
- Haag, M. and Willis, H. (2007). *Choctaw Language and Culture: Chahta Anumpa*, volume 2. University of Oklahoma Press.
- Haas, M. R. (1979). Southeastern languages. In Lyle Campbell et al., editors, *The Languages of Native America: Historical and Comparative Assessment*, pages 299–326. University of Texas Press.
- Leuski, A. and Traum, D. (2011). NPCEditor: Creating virtual human dialogue using information retrieval techniques. *AI Magazine*, 32(2):42–56.
- Li, X., Tracey, J., Grimes, S., and Strassel, S. (2016). Uzbek-English and Turkish-English morpheme alignment corpora. In *LREC*.
- Nicklas, T. D. (1972). *The Elements of Choctaw*. Ph.d. dissertation, University of Michigan.
- Nicklas, T. D. (1979). *Reference grammar to the Choctaw Language*. Choctaw Bilingual Education Program, Southeastern Oklahoma State University.

- Sturtevant, W. C. (2005). History of research on the native languages of the Southeast. In Heather K. Hardy et al., editors, *Native Languages of the Southeastern United States*, pages 8–65. U of Nebraska Press.
- Swartout, W., Traum, D., Artstein, R., Noren, D., Debevec, P., Bronnenkant, K., Williams, J., Leuski, A., Narayanan, S., Piepol, D., Lane, C., Morie, J., Aggarwal, P., Liewer, M., Chiang, J.-Y., Gerten, J., Chu, S., and White, K. (2010). Ada and Grace: Toward realistic and engaging virtual museum guides. In Jan Allbeck, et al., editors, *Intelligent Virtual Agents: 10th International Conference, IVA 2010, Philadelphia, PA, USA, September 20–22, 2010 Proceedings*, volume 6356 of *Lecture Notes in Artificial Intelligence*, pages 286–300, Heidelberg, September. Springer.
- The Choctaw Nation of Oklahoma Dictionary Committee. (2016). *Chahta Anumpa Tosholi Himona: New Choctaw Dictionary*. Choctaw Print Services, 1st edition.
- Traum, D., Jones, A., Hays, K., Maio, H., Alexander, O., Artstein, R., Debevec, P., Gainer, A., Georgila, K., Haase, K., et al. (2015). New Dimensions in Testimony: Digitally preserving a Holocaust survivor’s interactive storytelling. In Henrik Schoenau-Fog, et al., editors, *Interactive Storytelling: 8th International Conference on Interactive Digital Storytelling, ICIDS 2015, Copenhagen, Denmark, November 30–December 4, 2015, Proceedings*, volume 9445 of *Lecture Notes in Computer Science*, pages 269–281, Heidelberg, December. Springer.
- Ulrich, C. H. (1993). The glottal stop in Western Muskogean. *International Journal of American Linguistics*, 59(4):430–441.
- Williams, R. S. (1999). Referential tracking in Oklahoma Choctaw: Language obsolescence and attrition. *Anthropological Linguistics*, pages 54–74.