



NATIONAL SCIENCE FOUNDATION

NATIONAL ENDOWMENT FOR THE HUMANITIES



Documenting Endangered Languages

SAMPLE APPLICATION NARRATIVE

This sample narrative conforms to a past set of guidelines.

Please consult the guidelines for the current competition on NSF's website at:
<http://www.nsf.gov/pubs/2005/nsf05590/nsf05590.htm>

Language Contact and Variation: A Discourse-based Grammar of Monguor
Arienne M. Dwyer

NSF DEL Project Summary

This project will analyze annotated field data to prepare a grammar of Monguor, an unwritten endangered Mongolic language of northern Tibet. The end product will be a series of analytical modules published initially as a grammar for a linguistic audience and adapted for speaker community use, in consultation with Monguor colleagues. The data basis will be multi-tier time-linked grammatical annotations of a representative sample of 100 hours of high-quality audio and video recordings. Fieldwork and initial annotation were funded by the Volkswagen Foundation (2000–2005). These materials, recorded largely by native speakers, are being archived according to best-practice standards at the MPI-DOBES archive; a sample is also available on the EMELD site as a School of Best Practice Case Study.

After this initial documentation, producing analytical reference materials is the next logical step. In the current project, the data will be mined for basic morphological and syntactic information and queried for patterns in variation. The P.I. will write grammar module per month, making a short fieldwork trip to China midway through the project. During the final three months, the modules will be revised into chapters for publication as a linguistic reference work.

The immediate production of print reference works is critical. Using the nine UNESCO evaluative factors of endangerment, only one factor is favorable: the relatively large number of speakers (up to 50,000 out of a total population of nearly 200,000). All other factors point towards extinction (shrinking language domains, incomplete transmission, low prestige, and the absence of an orthography, media, and educational materials).

Intellectual Merit. This data-driven analysis hopes to make significant theoretical contributions in the syntax-pragmatics interface and in contact theory by applying corpus linguistics techniques to spontaneous spoken language data. Monguor's three dialects (North, Southeast, and Southwest) have had differing degrees of contact intensity with the dominant languages Tibetan and Chinese, allowing the evaluation of theories of grammaticalization and creolization. This holistic interdisciplinary study will be firmly anchored in the discourse and cultural context, examining how grammar emerges from discourse and interaction, particularly in the areas of argument structure, evidential and discourse marking, and intentionality.

Broader Impacts. Central Asia is poorly represented in the United States, despite its strategic importance and cultural diversity. This project strengthens a new Central Asian program at the University of Kansas, where the data will become available in the Linguistic Diversity Resource Center the P.I. is establishing, which allows students and interested researchers to conduct their own data queries. This project also includes women and Monguor researchers at its core, representing a politically powerless minority group in a remote area of China. The trilateral cooperation between China, Germany, and Kansas has resulted in a half-dozen faculty and student scientific exchanges and joint projects. The results would reach audiences in both China and the West: Monguor communities are enthusiastic about learning materials; these will heighten the status of the language for both Monguors and Chinese. Monguors have honed employable skills in media processing and translation. This project will encourage others to use a speaker-community-centered model of language documentation.

TABLE OF CONTENTS

For font size and page formatting specifications, see GPG section II.C.

	Total No. of Pages	Page No.* (Optional)*
Cover Sheet for Proposal to the National Science Foundation		
Project Summary (not to exceed 1 page)	1	_____
Table of Contents	1	_____
Project Description (Including Results from Prior NSF Support) (not to exceed 15 pages) (Exceed only if allowed by a specific program announcement/solicitation or if approved in advance by the appropriate NSF Assistant Director or designee)	13	_____
References Cited	4	_____
Biographical Sketches (Not to exceed 2 pages each)	2	_____
Budget (Plus up to 3 pages of budget justification)	3	_____
Current and Pending Support	1	_____
Facilities, Equipment and Other Resources	1	_____
Special Information/Supplementary Documentation	0	_____
Appendix (List below.) (Include only if allowed by a specific program announcement/ solicitation or if approved in advance by the appropriate NSF Assistant Director or designee)	_____	_____
Appendix Items:		

*Proposers may select any numbering mechanism for the proposal. The entire proposal however, must be paginated. Complete both columns only if the proposal is numbered consecutively.

NSF DEL

Language Contact and Variation: A Discourse-based Grammar of Monguor

Arienne M. Dwyer

Project Description

1. Introduction

This project aims to analyze annotated field data to prepare a discourse-based grammar of Monguor, an unwritten endangered Mongolic language complex of northern Tibet. The end product will be a series of analytical modules to be published initially as a grammar for a linguistic audience. Near the end of the grant period, some modules will be adapted of speaker community use, in consultation with Monguor colleagues.

The analysis will broaden the theoretical frameworks of discourse analysis and language contact by applying the techniques of corpus linguistics to spontaneous spoken language data. Monguor (Chinese *Tu*), spoken in Qinghai province in the People's Republic of China, comprises North, Southeast, and Southwest dialects. The Monguor language complex allows the evaluation of grammaticalization and creolization theories, as its three dialects have had differing degrees of contact intensity with the dominant languages Tibetan and Chinese.

This constitutes the start of the final phase of a multi-year documentation project whose fieldwork and initial annotation were funded by the Volkswagen Foundation (2000–2005). The relatively large number of speakers (ca. 25% of a population of over 120,000, which belies the rapidity of language loss) has allowed documentation of an unprecedentedly large number of communication genres in all three varieties. Though the Monguor recordings and minimal annotations will be publicly available via a web archive in five years' time (DOBES project archive 2001-present), analytic reference materials are still lacking. Producing a grammar is the next logical step in the documentation of Monguor.

During the grant period, these data will be mined for basic morphological and syntactic information and queried for patterns in variation. I will write one grammar module per month, making a short consultative fieldwork trip to Monguor areas during the ninth month of the project. During the final three months, these modules will be revised into chapters for publication as a linguistic reference work.

The strengths of this project are its data integrity, its modularity, and theoretical basis. The analysis is based on copious amounts of high-quality spontaneous spoken language data. From the Volkswagen Foundation sponsored project, *The Salar and Monguor Documentation Project*, we have a corpus of about 100 hours of audio and video data with multilayered transcriptions and analyses, which are being archived at the MPI archive in Nijmegen. Standards for the Monguor data are conformant with current best practices: metadata (OLAC, Simons and Bird 2002), character encoding (Unicode Standard 2004), annotation (DOBES archive 2001–2004, CES, Ide 1996), and text (TEI, Sperberg-McQueen and Burnard 1994).

The current grammar project is conceived of as a series of extensible modules, each corresponding to grammatical level or problem. For this grant period, the modules will constitute the basis for a linguistic reference grammar; thereafter, they will be used for a community reference grammar.

Applying techniques of discourse analysis to endangered-language data is relatively new, and will allow insight into typological universals of discourse. Combining this approach with language contact theory will give us a detailed picture of specific areal phenomena as well as advancing theories of contact-induced language change. In the last two decades, a number of excellent discourse-based studies have appeared (e.g. Givón 1983, Miller and Weinert 1998), but none have made extensive reference to data from endangered languages. This project would begin to bridge that gap.

2. Background and Groundwork

2.1. The Significance of Monguor

Figure 1: Sketch Map of Monguor areas



source: EMELD 2004.

Monguor (Mongolic) belongs to the greater *Sprachbund* of Amdo Tibet, today part of China's northwestern Qinghai province. The areas under investigation are largely in Qinghai province, with some Monguor speakers found in neighboring Gansu province.¹

Monguor is an unwritten language with rapidly diminishing fluent native speakers, particularly in the younger generation. Oral art forms central to both cultures are only known to the oldest generation; neither schooling nor media are available in these languages. Monguor is under intense pressure from the dominant languages, Chinese and Tibetan. Several plans by native speakers to develop orthographies and document their own languages have floundered due to a lack of equipment, computer skills, and publication possibilities.

These peripheral language varieties constitute keys to the study of Central Asian linguistic development by their preservation of archaic features of Mongolic. Monguor ethnogenesis likely began with their arrival in Amdo Tibet in the early 13th century as a part of the invading Mongolian army; certain oral accounts suggest an earlier migration from Mongolia proper. Thirteenth- and fourteenth-century Middle Mongolic phonological and lexical features in

¹Qinghai province counties with significant Monguor populations include Minhe (villages in all major Monguor townships), Tongren, and Huzhu (Danma, Wushi, Dongshan, Dongou, Sunduo, Donghe, and Taizi Townships and the Weiyuan Region). Monguors are also found in Ledu County (Dala Monguor Autonomous Township). In Gansu province, Monguors reside in Zhucha Township of Tianzhu County.

Monguor indicate that its speakers were geographically separated from the central Mongolian groups and settled in their present locale during this period.

Within what we here call “Monguor” (known in Chinese scientific literature as *Tu*), four main language varieties can be identified: Northern (Huzhu Mongghul), Southeast (Minhe Mangghuer), and Southwest (Tongren Mongghul). They are highly divergent, with variation is due to the intensity of language contact with Tibetan and Chinese. One variety has developed limited phonemic tonal contrasts due to contact with Chinese; all three varieties have Tibetan-style evidential systems. Current fieldwork suggests that the Southwestern Monguors were originally Tibetan populations who became “Mongolized” under contact with Mongolic speakers. The Monguors in the north and southeast are by contrast Tibeticized and Sinified Mongols, respectively.

2.2. Groundwork

Previous documentation on Monguor varieties is thin and uneven, and focuses primarily on the northern variety alone: a grammar (Mostaert and De Smedt 1929, 1930; De Smedt and Mostaert 1945), lexicon (Li 1988, Qinggertai 1986), and a few articles (Georg 2003, Dpal-Idan-bkra-shis et al. 1996). In addition, a book-length study has appeared on the Southeastern (Minhe) dialect (Slater 2003). Rich early ethnographic descriptions are available for the northern area (Schram 1954, 1957, 1961; Schröder 1959, 1970 and more recently from our team members e.g. Wang 2001, Zhu and Stuart 1996). For Southwestern (Tongren) Monguor, no published study exists at all.

As a result of our Volkswagen Foundation’s DOBES (Documentation of Endangered Languages) project, we have ample annotated and unannotated audiovisual material on all three Monguor language varieties. The basic fieldwork and annotation phases were supported by the program between October 2001 and May 2005. The aim as defined by the Volkswagen Foundation was to produce minimally annotated video and audio, archived in the Netherlands, for eventual Internet dissemination. The current project, however, shifts the focus to the production of analytical resources both for the linguist and the community. The first resource would be the grammar proposed here.

In the DOBES project, the research has covered historically important, lesser-known dialects, nearly extinct song forms and ethnomedical aspects of culture as well as linguistic features. Importantly, the project was speaker-centered, with a team of experienced and inexperienced local researchers conducting the bulk of the fieldwork and preliminary annotation.

Native speakers supplied their input on the important communicative genres for each area.² Our documentation team then used this list as a basis for fieldwork, recording two to ten samples of each genre. These high-quality digital audio and video recordings³ were captured, and catalogued with production and recording metadata,⁴ and archived on CD and DVD. Copies of these media were then returned to the documenters, who created a time-linked orthographic annotation and Chinese or Tibetan gloss of the audio stream in Unicode-compliant XML-based

² Sample communicative genres for Southwest Monguor are: conversation, festivals (5 types), child's one-month anniversary, son's party, weddings, funerals, stories (3 types), riddles, procedural texts, greetings and leave-takings.

³ Digital Audio Tape (DAT) recorded in stereo at 48 and 44.1 kHz, captured with a linear sound card at the same frequencies and archived in .wav format, digital video (PAL miniDV) archived as MP2.

⁴ Metadata is data about the event, its speakers, and the recording itself.

software (Bray et al. 2004). Phonetic transcriptions, English glosses, and were added later in the US and Germany; to about 15% of the corpus preliminary morphosyntactic annotation was added (both literal glossing and part-of-speech annotation).⁵

It is these richer time-linked annotations, which will constitute the data basis of the current project. A multi-tier annotation of a sample utterance from Southeastern Monguor is given in Figure 1 below. Currently, the corpus contains the following tiers: phonetic, orthographic, part-of-speech, morphological gloss, English gloss, and Chinese gloss. Annotation tiers are linked at the level of utterance by their time codes.

FIGURE 2. Annotation tiers (Source: DOBES Salar-Monguor Project, Dwyer, P.I.)

Session_tier	Utterance	Time_code
MNADDA23Jan0301_23_28_p	tsitɕʰaŋ awuni kə ʰpaŋ pa	109.53-110.13
MNADDA23Jan0301_23_28_o	Zichang awuni ger bang ba?	109.53-110.13
MNADDA23Jan0301_23_28_pos	Npr N-GEN N COP PRTconf	109.53-110.13
MNADDA23Jan0301_23_28_mge	Zichang paternal.younger.uncle-GEN home is PRTconf	109.53-110.13
MNADDA23Jan0301_23_28_ge	It's Uncle Zichang's family, isn't it?	109.53-110.13
MNADDA23Jan0301_23_28_gc	是子昌叔叔的家吧？	109.53-110.13

The available text samples coded in detail for morphosyntax, currently numbering ca. 20,000 words, will exceed 40,000 words by the beginning of the grant. These are a representative range of the relevant communicative genres of each area.

Prior to the NSF grant period, the grammatical annotation of the Monguor corpus will be internally standardized so as to be compliant with the guidelines of the Text Encoding Initiative (Sperberg-McQueen and Burnard 1994) and the morphosyntactic annotation recommendations of the Corpus Encoding Standard (Ide 1996, cf. also Leech et al. 1998). Designed to be extensible, these guidelines will allow this project to encode a non-Indo-European language like Monguor with attributes that are still conformant to emerging international guidelines for spoken language corpora.

Other by-products of the DOBES documentation of Monguor and Wutun will be employed in the current project. These include phonological and morphological sketches, morpheme tag set lists, preliminary lexicons, and fieldwork interview questionnaire results. Additionally, my databases from the mid-1990s of local Salar (Turkic) and Chinese will be employed in investigating areal phenomena.

3. Methodologies

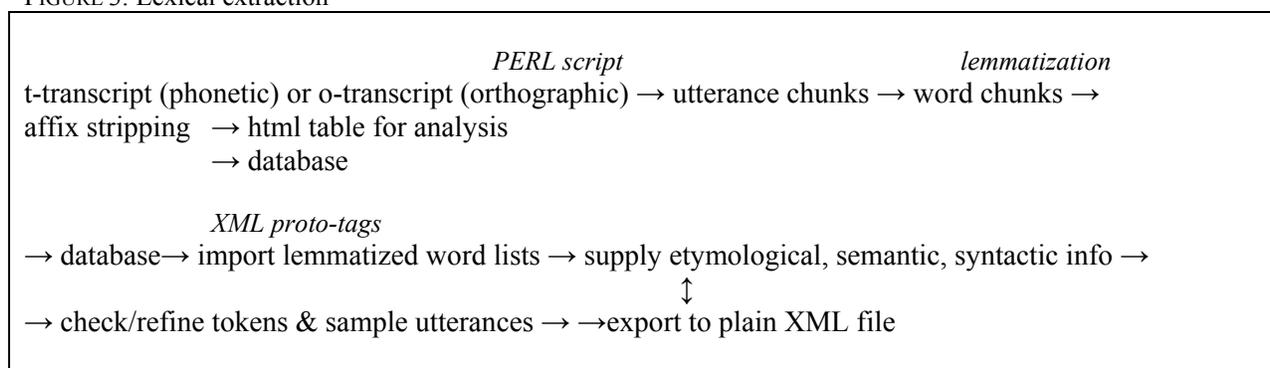
To complete a discourse-based grammar of Monguor, I will employ simple corpus linguistic methods to mine the Monguor data. These data will be the basis for the analysis, to be written concurrently. This mining - analyzing - writing cycle will be repeated at least nine times, for each module.

⁵ The workflow and data-conversion sequences of the Monguor project, as well as the native speaker centered research process is described in detail in the EMELD School of Best Practice site (EMELD 2004).

As I learn more about the grammar, I will also be refining the morphosyntactic annotations, likely adding a functional (i.e. grammatical role) tier (Pirelli and Sorria 2000) and a discourse tier (cf. Schmidt 2004) to supplement the two part-of-speech tiers for a small experimental subset of the data.

Data querying will entail simple string searches (e.g. for suffixes, particles, POS tags), key word in context (KWIC) concordancing, genre searches (wedding speeches, love songs, conversation), and some frequency analysis. Analysis of dialect variation will also be facilitated by lexical extraction. In its simpler form, tokens and/or lemmatized forms of the same feature (e.g. the locative case suffix) in all three dialects can be extracted and viewed in an html table for easy comparison. These forms will also be extracted into a database (exportable as a plain XML file, see Figure 3 below), so that a beta version of an electronic lexicon will also be created as a byproduct of annotated corpus analysis for this project.⁶

FIGURE 3. Lexical extraction



These querying methods - string and KWIC searches, concordancing, frequency analysis, and producing tables via lexical extraction - will allow me to compare the three dialects for each level of language, from phonetics to pragmatics. These features will be analyzed in nine modules.

This corpus-based approach should allow me to develop a detailed picture of the patterning of particular language-specific phenomena, as well as to test hypotheses about universals. Lexical extraction into tables should aid not only constituent descriptions but also investigations into discourse and evidentiality marking. As an example of the kinds of data I will be extracting, Figure 4 shows a partial list of discourse particles and their tentative POS tags, extracted from the XML annotation file. I added the tag glosses and commentary manually, and used the full extracted list to write an illustrated outline of the functions of Monguor discourse particles.

⁶ In the future we aim to produce a stand-alone multilingual lexicon as well as making the data and analyses available online through the Max Planck Institute's DOBES site.

FIGURE 4. Excerpt from extracted discourse particle (PRT) list

<i>particle</i>	<i>POS tag</i>	<i>gloss; comments</i>
ba	PRTconf	confirmative particle, sentential
ba	PRTsugg	suggestive particle (=hortative?); cf. Eng. "let's"
bei	PRTh	hortative particle
ge	PRThort	sentential hortative particle
ge	PRTsugg	suggestive particle, nonfinal cf. -gansige
-ge	PRTindef.sg	indef sing particle (<↑?), written tog w prev. in orthog, =clitic? need to stdze. ge also listed here as PRTh. = same morpheme??
lou	PRTdub	dubitative (concessive? VP1 lou Vp2)

The process of grammatical discovery is iterative and cyclical: these reflect the first-pass grammatical tags of the Monguor corpus, which are necessarily tentative. Armed with such a glossed list and a discourse sketch, one re-visits a larger portion of the corpus, refines the POS tags and the sketch, and thereafter can produce a draft module. Discourse particles are but one grammatical feature under investigation; verb-serializing, evidential, and clause-chaining suffixes and other phenomena would all benefit from lexical extraction and repeated analysis.

No matter how much excellent data one has, one inevitably finds gaps in the grammatical description that cannot be filled by the current data. The ideal solution is a brief follow-up fieldwork trip, which this project has planned two-thirds of the way through. I intend to go with a long list of questions and a draft of the grammar modules for consultation with Monguor colleagues. Thereafter, my principle activity will be revising the analysis and preparing the nine modules for publication as a monograph.

These methods conform to the best possible practices of the current technologies and state of linguistic research in the sense that the analysis is data-driven, that is, it is based on spontaneous spoken data. The difficult issue of language variation is not glossed over, but rather made central to the investigation. I make judicious use of discourse analysis and other theories as the basis for the grammar, which ensures that I am not imposing *a priori* categories on a poorly understood language. Finally, this project can be considered to be "best-practice" in that encoding and annotation standards for the Monguor data are OLAC, EMELD, and MPI-DOBES-conformant.

4. Work schedule

15 August 2005 - 14 August 2006

Every month I will analyze a new topic in Monguor grammar, beginning with the language contact setting and the sound system, and working through morphology and syntax to specific topics in discourse. Annotated data will be queried for these features concurrently to the analytical writing. The result will be one analytical grammar module per month. As the grammatical analysis is refined, a list of outstanding research questions will be drawn up for checking in Monguor areas during the ninth month. During the final three months, these modules will be revised for publication. New evidence brought to light by grammatical analysis may result in a slight change in focus from the modules listed here to the final book chapters.

Some goals as outlined in Figure 5 below require explanation:

Month 1: Preparing the data set for analysis

At the outset, it is critical to evaluate and refine the morphosyntactic annotation of the prior DOBES project. Specifically, internal consistency and compliance with international annotation standards must be assessed before grammatical analysis can begin. This will be accomplished by checking the part-of-speech (POS) annotation for consistency between the three dialect groups. These markup "tags" will then be mapped onto the GOLD ontology (Farrar, Lewis, and Langedoen 2002, EMELD 2003), a process our team began in 2004 with the EMELD FIELD tool (EMELD 2004).

Module 7: Revisiting the comparative-historical question of Wutun and Monguor

Wutun is a Tibetan-influenced Chinese language variety with Mongolic components spoken in the same valley as Southwestern Monguor in Tongren County.⁷ Our DOBES project included the documentation of Wutun. Though the grammar of Wutun deserves fuller treatment on a separate later grant, in this current project it is still essential to compare the Mongolic varieties of Tongren County. A comparative analysis will clarify historical questions and allow us to better assess the causes and degree of variation between Southwestern Monguor and the other varieties of Monguor.

Module 9: Language Contact Phenomena: Areal Features

The Monguor varieties are located at the center of the north Tibetan *Sprachbund*, and share a number of features with the other languages of the area. Such features contact-induced features include retroflex/palatal initial obstruent series contrasts, lexicon, and evidentiality.

⁷ Wutun is an under-investigated language variety (Chen 1982, 1986; Tongren...1983) that has gained attention in the West due to the heavy Chinese-Tibetan-Mongolic contact phenomena that it exhibits (Thomason & Kaufman 1988, Li 1983). It is one of only eight so-called "mixed languages" recognized in the Ethnologue (http://www.ethnologue.com/show_language.asp?code=WUH).

FIGURE 5. Work schedule

Month	Goal	Tasks
9/05	Evaluate data set's readiness for analysis; Prepare Module 1: overview of the Monguor language and contact environment	-Regularize POS tags across Mng. dialects; -Map POS tags to GOLD ontology; -Generate a list of unique words; -Refine lexical extraction routine; -Draft Module 1
10/05	Module 2: Phonetics & Phonology	Analysis of the sound systems, using existing and new transcriptions; -Transcribe new phonetic samples of all three dialects -Generate phonetic-phonemic tables via Perl script
11/05	Module 3: Nominal & Verbal Morphology	-POS searches -Lexical extraction
12/05	Module 4: the Clause and Clause combining; Syntax	-Experiment with functional markup -Bracketing and dependency relations -Concordancing
1/06	Module 5: Discourse Pragmatics and TAM marking	-Genre-specific searches -Frequency counts
2/06	Module 6: Evidentiality	-KWIC searches for evidential suffixes and adverbs; -Comparison with Tibetan system
3/06	Module 7: Revisiting the comparative-historical question of Wutun and Monguor	-Generate comparative tables: Wutun vs. SW Monguor -Comparative analysis of Wutun morphosyntax
4/06	Module 8: Discourse Marking	-Particles search; classification -Anaphora (tracking e.g. PrNs + persPN reference)
5/06	Fieldwork: check hypotheses with speakers (in China)	Consultative discussion and field work in N. Tibet; fill in gaps in analysis; assess community grammar needs
6/06	Make Changes & Additions	-Incorporate fieldwork data into previous analyses -Begin revising modules for publication
7/06	Module 9: Language Contact Phenomena: Areal Features	-Compare NW Chinese, Amdo Tibetan, and Salar databases
8/06	Revise modules for publication	

Though compiling a reference grammar of three divergent dialects of an endangered language is a big task, the groundwork of data collection, annotation, and preliminary analysis has already been accomplished. All of our team's previous work on Monguor (2001–2005) and my work and databases on its contact languages (1991–2000) will facilitate data analysis and completion of the grammar.

5. Participants: roles and qualifications

The qualifications necessary for this Monguor grammar project are primarily computational resource management skills and experience in the linguistic analysis of language contact and variation. A thorough knowledge of the typologies of Mongolic and Monguor's primary contact languages (varieties of Chinese and Tibetan) is also highly desirable.

As PI of the current project, I have eighteen years of field experience in Northwest China, having spent a total of forty-three months in the field working on contact and variation in Sinitic, Turkic, and Mongolic languages. I speak the major *linguae francae* of the area (Mandarin

Chinese, Uyghur, and am learning Amdo Tibetan) and have minimal to intermediate competence in all the language varieties I have investigated, including Monguor. Earlier I conducted linguistic fieldwork on Northwestern Chinese syntax (Dwyer 1995), language contact phenomena in modern Uyghur dialects (Dwyer 1999, 2001a, 2004), Kazakh (Dwyer 2001b), and especially on Salar grammar (Dwyer 1998, 2000a, 2000b, 2005 forthcoming), creating database systems for text and audio data for these language varieties.

Many of the skills I honed during research on Salar (Turkic) have been directly applicable to the Monguor varieties, since the language contact profile of Salar closely resembles that of Monguor, given the typological similarity and geographical proximity of Salar.

Most recently I have directed a team of twenty-two people (in China, Germany, and the United States) in the VW-DOBES sponsored *Documentation of Salar and Monguor* project, focusing on Monguor and Wutun. In China, this entailed building a research infrastructure in northern Tibet (with field and data-entry equipment in the three relevant counties and a digitizing center and language archive in the provincial capital) and training a dozen Monguors in basic field and orthographic annotation techniques. In Germany and the U.S., I oversaw the standardization and further annotation of the Monguor resources. The Salar/Monguor team has played a central role during the pilot year in developing annotation and metadata standards for the DOBES consortium

Though not directly involved in the current project, it is important to briefly introduce the Monguor research team in northern Tibet, given their central role in data collection and processing. The local management team consists of A, B, and C. A is a native speaker of Minhe Monguor and a medical doctor. He has coordinated the project in China since 2000 and has contributed his own folkloristic and lexicographic research, on which he has a number of published articles (e.g. Dpal-ldan-bkra-shis et al. 1996, Wang 2001). B has had extensive experience collecting, recording, and transcribing folklore. He is a fluent Northern Monguor speaker. C, also a Northern speaker, has made videos and transcriptions of Huzhu Monguor for the past five years and overseen the a/v capturing and archiving in situ.

In addition, we have three local dialect area teams who have now three years of experience making recordings and doing basic annotation: the Minhe team , the Huzhu team , and the Tongren team .

6. Degree of Endangerment

What is unusual about Monguor as an endangered language is the relatively large number of speakers (up to 50,000 out of a total population of nearly 200,000). But other factors point to a rapid decline in all varieties: language domains are shrinking and neither an orthography nor educational materials are available in the any form of Monguor, so that the language is already being incompletely transmitted between generations (see Figure 6 below). Though a few materials are available in a locally adapted orthography (e.g. Li 1988), these are not in wide circulation, and the orthography is not officially sanctioned. As a result, no print media exist in Monguor, and other media (TV, film, video, radio) are also not available in the language. The language is not taught in schools at any level, which reinforces the low prestige of the language varieties in the eyes of speakers, many of whom see Monguor as "useless" for modern life. The language is also low prestige for the Chinese government, as indicated by its unwillingness to approve an orthography and schooling in the language. Most importantly for the proposed

project, the coverage, quality, and amount of linguistic documentation is currently inadequate: what sources exist describe only the flagship Northern dialect, ignoring the others. Quite recently, an excellent comprehensive study of the Southeastern (Minhe) dialect has appeared (Slater 2003), but only one discourse genre is represented and no recordings are available. The Southwestern dialect, though significantly different, has only a mimeographed study (Tongren 1983).

- *Rapid loss of native speakers*

The number of fully fluent speakers is decreasing rapidly, with only 15-30% of the population having an active command of the language.⁸ Generational differences are particularly acute in all areas: though speakers over 60, particularly women, are fully fluent and competent in oral art forms and are native-language-dominant, speakers over 35 have a passable-to-fluent command of their native language, but are multilingual and have no command of oral art forms, whereas children may grow up with one of the dominant languages as their native language.

Even as the language rapidly changes, the large number of current speakers has allowed our VW-DOBES project the chance to do an unusually *complete* documentation, with nearly 100 hours of audio and video recordings representing a dozen communicative genres in each of the four areas.

- *Poverty and lack of native-language education*

Those areas where Monguor language and folklore are best preserved are characterized by remoteness, extreme poverty, and lack of education, for example villages in Huzhu and Minhe counties. The lack of native-language schooling and a writing system for Monguor children makes the future of the language bleak.

- *Forced resettlement will result in language death*

In late 2001, the provincial government approved a dam-construction project at a point on the Yellow River just at the Minhe Monguor areas. Already toward the end of this research project (2006), ca. 5000–7000 Monguors will be forced to abandon their villages and resettle in distant Haixi prefecture, ca. 1000 km to the west. Being cut off from the already linguistically fragile littoral communities will almost certainly result in language death within two generations for these speakers.

- *Summary: Nine Endangerment Criteria*

Using the UNESCO Intangible Cultural Heritage Unit's nine factors to assess the overall endangerment situation of Monguor, it becomes apparent how critical the need is for print reference materials.

FIGURE 6. Degree of endangerment: Monguor

⁸We define fluency here as not just conversational fluency, but also competence in other oral forms central to culture: storytelling, oral history, singing, oratory, ritual. Estimates based on fieldwork by Dwyer in 2001–2003.

(UNESCO 2003; 0=absent/extinct, 3=definitively endangered, and 5=not endangered)

Factor / Grade (0-5)	Monguor (MJG)		
	SE (Minhe)	N (Huzhu)	SW (Tongren)
1. Intergenerational Language Transmission	3	4	4
2. Absolute Number of Speakers (est.)	3750	45,000	625
3. Proportion of Speakers within the Total Population (estimated)	15%	30%	25%
4. Loss of Existing Language Domains	2	3	3
5. Response to New Domains and Media	0	0	0
6. Materials for Language Education and Literacy	0	1	0
7. Language Attitudes and Policies	2	3	2
8. Community Members' Attitudes toward Their Own Language	2	2	2
9. Amount and Quality of Documentation	2	2	0-1

This project directly targets the endangerment factors with the lowest ratings (in **bold** above). Our annotated language resources constitute new media, and the Monguor grammar will serve as the foundation for a community reference grammar, i.e. language education materials. Outside of our DOBES materials, no A/V is currently available anywhere. But, as Factors 5, 6, and 9 show, materials in print such as a reference grammar are equally if not more urgently needed.

7. Significance of the Project: Intellectual merit

In producing reference materials on endangered languages, we must set standards that are as high as for any other field of linguistics. Endangered-language linguistics is doubly ghettoized, once because the speaker communities are usually politically unimportant, and twice because in at least American linguistics today descriptive reference works are lower prestige than theoretical works. But this project is attempting to reverse this trend, by involving the speaker community to the greatest extent possible, and by addressing theoretical issues directly in a descriptive reference work.

The evaluation of issues in discourse and language contact theory is only feasible with a significant body of data such as the present Monguor corpus. In discourse theory, the contribution I hope to make is at the syntax-pragmatics interface. Such polarities were thematized by Givón (1979) but brought into particularly sharp focus with the emergent grammar hypothesis of Hopper (1987). This study will examine how grammar emerges from discourse and interaction, particularly in the areas of argument structure, evidential marking, and discourse particles. Speaker intentionality plays an important role in this area (cf. Fig. 4 above). This Monguor grammar, then, aims to be firmly anchored in the discourse context.

In the field of language contact studies, this project plans to evaluate each level of language for contact-induced and other language change in the three Monguor dialects. Predictability is one issue: while specific changes are likely not foreseeable, the probability of language decline based on current contact-induced change may be indeed predictable. Another issue is the potential correlation between typological similarity among contact languages and degree of language change. (There is no such correlation in northern Tibet.)

While the proposed discourse analysis of Monguor is resolutely synchronic, the language contact investigation will make reference to diachrony. Grammaticalization studies (e.g. Hopper and Traugott 1993) have shown that it is often relevant for synchrony to make reference to diachrony.

The holistic documentation of endangered languages requires drawing on many academic disciplines and subdisciplines beyond linguistics, including information technology, art history, sociology, ethnobotany, and ethnomusicology. One focal point of the Monguor documentation project was healing systems. Without knowing the significance of the objects or clothing associated with shamanizing or Taoist fortune-telling, the ritual language of healing ceremonies is much more difficult to interpret. Another focal point has been Monguor song forms, especially love songs. From this research I have hypothesized that musical ornamentation is copied rapidly between dominant and non-dominant groups, analogously to phonological copying (a.k.a. "borrowing"). Through this interdisciplinary investigation, our team has come to view discourse and language contact as phenomena emerging from the language within the broader cultural context.

In short, this grammar will be a data-driven linguistic study that hopes to make significant theoretical contributions in the syntax-pragmatics interface and in contact theory.

8. Broader Impacts

In the United States, Central Asian Studies has been poorly represented, despite the region's strategic importance and cultural diversity. This project strengthens a budding small Central Asian program at the University of Kansas, where the data and refinements in querying capabilities will be available in the *Cultural and Linguistic Diversity Resource Center* I am in the process of establishing. The enhanced morphosyntactic annotation of this project will allow students and other interested researchers to conduct their own data queries.

This possibility of hands-on corpus analysis assignments will enhance student learning in e.g. my *Linguistic Data Processing* class. The Monguor grammar will also contribute to our understanding of the Mongolic language family within the context of Central Asia.

In terms of enhancing the participation of underrepresented groups, this project's research is based on the fieldwork and preliminary annotations of Monguor colleagues and assistants, who represent one of China's more obscure and politically powerless minority groups in a remote area. I have gone out of my way to employ women there.

This cooperation between citizens and institutions in western China, Germany, and Kansas has already resulted in a number of scientific exchanges and joint projects: a German computational linguist and a Monguor academic have both lectured and had research stays at the University of Kansas, and the latter is planning to send its first students to Inner Asia next autumn. Support for this project would send a strong signal of support for further cooperation, particularly at a time when the paperwork for international exchanges is increasingly difficult.

The results of this research should reach the broadest possible audience with three planned formats: first to be published in the West as a linguistic reference work, then reworked as a community grammar and published in Chinese in China, and finally Web-based and CD-ROM dissemination is planned.

The societal impacts of producing a Monguor grammar would be significant both in China and the West. All of the Monguor communities we have visited and lived in expressed

unbridled enthusiasm for the production of any sort of learning materials, video, and/or audio recordings. Creating a linguistic reference grammar, which will be the basis for a later community grammar, will heighten the status of the language for both Monguors and Chinese in a society where the written word has enormous prestige. Continued work on this project would also enable the dozen Monguors employed on this project to hone their employable skills in computer, translation, and media processing.

For the general public in the West, such a grammar could serve to replace the stereotype of the 'bloodthirsty conquering Mongols' with a much more nuanced view of this Tibetan Buddhist people.

Supporting this grammar will hopefully encourage more projects to use a speaker-community-centered model of language documentation. For the last four years, a principal organizing feature of our project has been that it is run by native speakers themselves. My Monguor colleagues direct the conceptualization and carrying out of our joint community-based project, whereas my role is largely that of a facilitator, providing equipment and training in the initial stages, and providing linguistic guidance and analyses in later stages. Thus it was Monguors themselves who decided which communicative genres should be investigated in each dialect area and who made most of the audio-video recordings.

In my view, a documentation is not complete until useful products for several audiences appear. For both academics and for communities, tangible products such as this grammar are equally important as annotated audio-video material. This is also an important step towards giving something substantial back to the community, in the form of a future community reference grammar.

NSF DEL

Language Contact and Variation: A Discourse-based Grammar of Monguor

Arienne M. Dwyer

References

- Bray, Tim, et al., eds. 2004. Extensible Markup Language (XML) 1.0 (3rd edition). W3c recommendations. <http://www.w3.org/TR/REC-xml>
- Chen Naixiong. 1982. Wutunhua chutan [A Preliminary Exploration of the Wutun vernacular]. *Minzu yuwen* 1: 10–18.
- Chen Naixiong. 1986. Guanyu Wutunhua/An Outline of the Wutun Linguistic Structure. *Journal of Asian and African Studies* 31: 33–52.
- De Smedt, A. and A. Mostaert. 1945. *Le Dialecte Monguor. II^e Partie: Grammaire*. Monumenta Serica 6, Journal of Oriental Studies. Peking: Catholic University.
- DOBES project archive 2001-present. <http://www.mpi.nl/DOBES>
- Dpal-Idan-bkra-shis, Hu Jun, Huping, Limusishiden (Li Dechun), Keith Slater, Kevin Stuart, Wang Xianzheng, and Zhu Yongzheng. 1996. Language Materials of China's Monguor Minority: Huzhu Mongghul and Minhe Mangghuer. *Sino-Platonic Papers*.
- Dwyer, Arienne M. 1995. From the Northwestern Chinese *Sprachbund*: Xúnhuà Chinese Syntactic Data. *The Yuen Ren Society Treasury of Chinese Dialect Data Vol I*: 143–182.
- 1996. When orthographies are verboten: Endangered-language policy in northwestern China. In Garland D. Bills, ed. *Southwest Journal of Linguistics Special issue: Language Loss and Public Policy II* :15:1/2.
- 1998. The Turkic Stratigraphy of Salar: An Oghuz in Chagatay Clothes? *Turkic Languages* 2.1: 49–83.
- 1999. Language Contact in Qumul [Xinjiang]. *Journal of Central Asia* special issue.
- 2000a. Direct and Indirect Experience in Salar. In Bo Utas and Lars Johanson, eds. *Evidentials. Turkic, Iranian and neighbouring languages*. Berlin: Mouton de Gruyter.
- 2000b. Consonantalization and Obfuscation. In Goksel, Asli & Celia Kerslake (eds.) *Studies on Turkish and Turkic Languages. Proceedings of the Ninth International Conference on Turkish Linguistics Lincoln College, Oxford, 12-14 August 1998* (Turcologica 46). Wiesbaden: Harrassowitz.

- 2001a. Uyghur. In Garry, Jane, and Carl Rubino, eds. *Facts About the World's Languages: An Encyclopedia of the World's Major Languages, Past and Present*. New York/Dublin: HW Wilson Press.
- 2001b. Dialogic Discourse in Performance. Unpublished manuscript.
- 2004. The Xinjiang Conflict: Uyghur Identity, Language Policy, and Political Discourse. *Policy Studies* 12. Washington, D.C.: East-West Center Washington.
- 2005 forthcoming. *Salar: a study in Inner Asian areal contact processes, Part I: Phonology*. Wiesbaden: Otto Harrassowitz.
- E-MELD. 2004. *E-MELD School of Best Practice: Story of the Monguor Data*. Online: <http://cf.linguistlist.org/cfdocs/emeld/school/exhibits/monguor/storyofdata.html#future>
- EMELD 2003. General Ontology for Linguistic Description. Online: <http://cf.linguistlist.org/cfdocs/emeld/school/ontology-tree.cfm>
- Farrar, S., Lewis, W. & Langendoen, T. 2002. *A Common Ontology for Linguistic Concepts*. Proceedings of the Knowledge Technologies Conference, Seattle, WA. <http://emeld.douglass.arizona.edu:8080/papers/KnowTech-CommonOntology.pdf>
- Georg, Stephan. 2003. *Mongghul*. in Juha Janhunen, ed. *The Mongolic Languages*. NY: Routledge.
- Givón, T. 1979. *On Understanding Grammar*. New York: Academic Press.
- Givón, Talmy. 1983. *Topic continuity in discourse: A quantitative cross-language study*. Amsterdam: John Benjamins.
- Hopper, Paul. 1987. Emergent Grammar. *BLS* 13: 139-157.
- Hopper, Paul, and Elizabeth Closs Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge UP.
- Ide, Nancy, ed. 1996 CES. The Corpus Encoding Standard. Online: <http://www.cs.vassar.edu/~ide/CES/CES1.html>
- Leech, G., Weisser, M. and Wilson, A.. 1998. Draft chapter: Survey and guidelines for the representation and annotation of dialogue. LE-EAGLES-WP4-3.1, Integrated Resources Working Group, 17 April. Online: <http://www.ilc.pi.cnr.it/EAGLES/home.html>
- Li, Charles. 1983. Languages in contact in western China. *Papers in East Asian Languages* 1: 31-51.

- Li Keyu. 1988. *Mongghul Qidar Merlong/Tu-Han cidian* [A [Huzhu] Monguor-Chinese dictionary]. Xining: Qinghai renmin.
- Miller, Jim and Regina Weinert. 1998. *Spontaneous Spoken Discourse: Syntax and Discourse*. Oxford: Clarendon Press.
- Mostaert, A. and A. de Smedt. 1929. Le Dialect Monguor parlé par les Mongols Kansu Occidental. partie 1^{ère} Partie: Phonétique. *Anthropos* XXIV: 145–165, 801–815.
- 1930. Le Dialect Monguor parlé par les Mongols Kansu Occidental. partie 1^{ère} Partie: Phonétique. *Anthropos* XXV: 657–669, 961–973.
- Pirelli, Vito and Claudia Sorria. 2000. MATE [Multilevel Annotation Tools Engineering] Level Markup: Functional Annotation Coding Module. Online: <http://www.ims.uni-stuttgart.de/projekte/mate/mdag/ms/Funct.htm>
- Qinggeertai [Chinggeltai], ed. 1986. *Tuzuyu Huayu cailiao* [Monguor language materials]. Hohhot: Nei Menggu Minzu.
- Schmidt, Thomas. 2004. *eXMARaLDA: Extensible Extensible Markup Language for Discourse Annotation*. Online: <http://www.rrz.uni-hamburg.de/exmaralda/>
- Schram, Louis M. J. 1954. *The Monguors of the Kansu-Tibetan Frontier: Their origin, history, and social organization*. Philadelphia: Transactions of the Americal Philosophical Society Vol. 44:1.
- 1957. *The Monguors of the Kansu-Tibetan Frontier: Part II. Their religious life*. Philadelphia: Transactions of the Americal Philosophical Society Vol. 47:1.
- 1961. *The Monguors of the Kansu-Tibetan Frontier: Part III. Records of the Monguor Clans*. Philadelphia: Transactions of the Americal Philosophical Society Vol. 51:3.
- Schröder, Dominik. 1959. *Aus der Volksdichtung der Monguor, 1. Teil*. Asiatische Forschung 6. Wiesbaden: Harrassowitz.
- 1970. *Aus der Volksdichtung der Monguor, 2. Teil*. Asiatische Forschung 31. Wiesbaden: Harrassowitz.
- Simons, Gary, and Steven Bird. 2002. Recommended Metadata Extensions. Open Language Archives Community, <http://www.language-archives.org/REC/olac-extensions.html>
- Slater, Keith W. 2003. *A Grammar of Mangghuer: a Mongolic Language of China's Qinghai-Gansu Sprachbund*. NY: Routledge/Curzon.

Sperberg-McQueen, C.M. and Burnard, Lou, eds. 1994. *Guidelines for Electronic Text Encoding and Interchange. TEI P3. Text Encoding Initiative*. ACH, ACL, ALLC, Chicago, Oxford. Online: <http://etext.virginia.edu/TEI.html>

Thomason, Sarah Grey and Terrence Kaufman. 1988. *Language Contact, Creolization and Genetic Linguistics*. Berkeley: UC Press.

Tongren Tuzu kaochazu. 1983. *Tongren Tuzu kaocha baogao: sizhaizi (Wutun) de minzu lishi, yuyan he yishu. Di er bu: yuyan*. [A Investigative Report on the Tu (Monguor) of Tongren: The four villages' (Wutun) nationality history, language, and art]. Xining: Qinghai Nationalities College, Nationalities Institute mimeograph.

UNESCO Expert Committee on Endangered Languages. 2003. *Language Vitality and Endangerment*. Paris: UNESCO. Online: http://portal.unesco.org/culture/en/file_download.php/1a41d53cf46e10710298d314450b97dfLanguage+Vitality.doc

Unicode Standard. 2004. <http://www.unicode.org/versions/Unicode4.0.1/>

Wang Xianzhen, with Zhu Yongzhong and Kevin Stuart. 2001. *Mangguerla Bihuang Keli* [Mangghuer Folktale Reader]. Xining: The Bridge Fund (internal circulation publication).

Zhu Yongzhong and Kevin Stuart. 1996. Minhe Monguor *Nadun* Texts. *CHIME* 9: 89–105.