

# Language Documentation and Description

ISSN 1740-6234

---

This article appears in: *Language Documentation and Description, vol 4*. Editor: Peter K. Austin

## The importance and challenges of documenting pragmatics

LENORE GRENOBLE

Cite this article: Lenore Grenoble (2007). The importance and challenges of documenting pragmatics. In Peter K. Austin (ed.) *Language Documentation and Description, vol 4*. London: SOAS. pp. 145-162

Link to this article: <http://www.elpublishing.org/PID/054>

This electronic version first published: July 2014

---



This article is published under a Creative Commons License CC-BY-NC (Attribution-NonCommercial). The licence permits users to use, reproduce, disseminate or display the article provided that the author is attributed as the original creator and that the reuse is restricted to non-commercial purposes i.e. research or educational use. See <http://creativecommons.org/licenses/by-nc/4.0/>

---

## EL Publishing

For more EL Publishing articles and services:

Website:	<a href="http://www.elpublishing.org">http://www.elpublishing.org</a>
Terms of use:	<a href="http://www.elpublishing.org/terms">http://www.elpublishing.org/terms</a>
Submissions:	<a href="http://www.elpublishing.org/submissions">http://www.elpublishing.org/submissions</a>

# The importance and challenges of documenting pragmatics

Lenore A. Grenoble

## 1. Introduction

The importance of documenting pragmatics for endangered languages would hardly seem to be controversial or even in need of explanation. To the extent that language documentation aims at a truly comprehensive record of a language, that record must necessarily include the functioning of language as situated in context. Pragmatic knowledge and linguistic use are fundamental parts of the ethnography of communication and are key to understanding basic communicative practices. Documenting a language is paramount to documenting culture: culture is indexed through language situated in use, and understanding cultural concepts requires an understanding of the discursive interactions which invoke it (Silverstein 2004). Hill (2006) makes a similar point, arguing for documentation which relies on the traditions of the ethnography of language and not linguistic analysis which divorces language from its context.

While this would seem to be an obvious, it is difficult to achieve. Documenting spontaneous, contextualised language use requires long access to the community, including access to a wide variety of situations in which language is used. The analysis of pragmatics demands a deep understanding of the language and the culture in which it is embedded; this is not something that is quickly achieved by an outside linguist. The basic collection, transcription and analysis of word lists, folk texts and other narratives, is itself sufficiently time-consuming that many linguists leave aside questions of pragmatic usage. Moreover, the study of pragmatics requires access to face-to-face, everyday interaction; such situations have been argued to be the 'primary' sources of information for research on the functioning of language, which is central to pragmatics (Mey 1993:48); in situations of extreme endangerment with very few fluent speakers, such face-to-face interactions may not occur at all. The study of pragmatics requires a speaker community and attention to extralinguistic features, including such things as pauses, the use of eye gaze, gestures, and so on, all of which can affect the illocutionary force of an utterance and are a vital part of language and culture.

Even when this is attainable, it is difficult to know at what point one has sufficiently documented the pragmatics of a given language. A useful and

diverse corpus should include a great variety of different discourse styles, registers and genres, along with analysis of when they are appropriately used. One way to judge what kind of documentation will be needed should a speaker community cease to be available is to consider existing documentation, with an eye toward what it does and does not say about pragmatic usage. In this paper I examine the potential for documentation to reconstruct pragmatic and situational uses of language through the prism of existing language description, with the documentation of one of the Tungusic languages, Evenki, serving as a case study. The purpose of the discussion is twofold. One is to explore the best practices for documenting and describing discourse-pragmatics. The other is to consider the linguistic consequences of language shift in contact situations and how pragmatics are affected (or not) by this shift. The documentation and description of pragmatics and discourse in endangered languages, in particular in those languages already undergoing shift, present certain difficulties. First of all, in shift situations, speakers are not always fully fluent and may have limited access to certain registers, genres and/or styles. In addition, because of the shift, their intuitions are not always reliable (or useful) and they may show interference from the contact language(s). These are obstacles which linguists face in any aspect of documenting and describing a language undergoing shift but pose particular challenges in the elicitation and analysis of pragmatics

## 2. Pragmatics and Tungusic languages

The Tungusic languages represent one of several branches of Altaic and are spoken in northern China and Siberia. The languages of this family are all endangered, to varying degrees, and predictions are that the entire family will be lost within the next fifty years or so (Janhunen 2005, Whaley 2003). The Tungusic languages would appear to be reasonably well-documented.<sup>1</sup> Manchu, the only member of the family to have a written form before the 20th century, has a large corpus of texts thanks to its historical political prominence, a written form of Manchu was developed in the 16th century and a relatively large corpus of texts of a variety of genres is extant (see Gorelova 2002). The remaining Tungusic languages do not have a long-standing written tradition. The Siberian Tungusic languages initially became the subject of study by Russian explorers of the late 19th century, and were more thoroughly studied by linguists as part of the Soviet campaign in the North; written forms were created for most of them in the 1930's by these linguists. Not all Siberian Tungusic languages have been studied to the same degree, but those with larger numbers of speakers (Evenki, Even, Nanai) and even some of

---

<sup>1</sup> For a bibliography see: <http://www.dartmouth.edu/~trg/biblio.html>

those with very few speakers (e.g. Oroch, Udihe) have been described in dictionaries, descriptive grammars and text corpora. There are some early recordings of these languages housed in St. Petersburg (see Burykin et al. 2005). The result is that we have a fair amount of material relating to the languages as they were spoken in the early 20th century but relatively few texts. Those which were recorded are primarily folktales and songs, while other genres were largely ignored. It is important to note that although written languages were created for some Tungusic languages as part of the Soviet literacy campaign, they did not achieve a bona fide level of usage. The Tungusic cultures, with the exception of Manchu, have remained primarily oral. The published description does not include texts for genres other than folklore, epics and songs. There is no conversational data, personal histories or descriptions of activities fundamental to Siberian Tungusic life and culture, such as reindeer herding, hunting and fishing, clan structure, and so on. They are limited in scope and content.

In sum, the existing description for the Tungusic languages consists of dictionaries, descriptive grammars and grammatical sketches, and some texts, varying in quantity among the different languages but with a variety of discourse types for only Manchu. One way to test the viability of such description is to see what information it can provide for further analysis. A full description should enable linguistic analysis at all levels. In 1979 Johanna Nichols published a paper entitled the “Syntax and pragmatics in Manchu-Tungus languages;” this is, to the best of my knowledge, the only published description of Tungusic pragmatics to date. This research was done during the Soviet era, at a time when it was extraordinarily difficult, and usually impossible, for an American linguist to do fieldwork in Siberia. Thus Nichols was forced to rely on the existing, published description of Tungusic. In essence her study replicates the situation which would occur if the Tungusic language family were to be lost; it is in fact a fair trial for the documentation.

We can at this point ask two fundamental questions. First, was the description Nichols had available adequate for the study of Tungusic pragmatics? Second, given her analysis and subsequent documentation, can we adequately describe the use of the Tungusic languages? Rather than address these questions for all Tungusic languages, in what follows I focus primarily on the pragmatics of Siberian Evenki, a Northwest Tungusic language spoken by an estimated 9000 speakers, some of whom are settled and live in separate villages, others continue to maintain a subsistence hunter-herder lifestyle and live at large in the taiga and tundra regions of Siberia. Contrastive information is considered from Oroqen, a closely related variety of Tungusic, spoken in China. Oroqen is so closely related to Evenki that many Russian scholars consider it to be a dialect of Evenki, although I have argued elsewhere (Whaley et al. 1999) that the differences between the two

are more language-like than dialect-like. Their exact status is, however, irrelevant our purposes here; the fact that there is dispute about their status is indicative of their very close relationship. Yet despite their genetic affinity, there are significant differences in word order and discourse structure between the two varieties. Current documentation makes it difficult to evaluate to what extent these differences are the result of areal influences, or of attrition, or of language-internal change. Evenki is in close contact with Russian, Oroqen with Mandarin; both are also in extensive contact with neighbouring Turkic and Mongolic languages. Both are also endangered and undergoing shift, primarily to Russian in the case of Evenki, and to Mandarin in the case of Oroqen.

Nichols presents a remarkably subtle and sophisticated analysis based on the relatively limited description with findings which have implications far beyond the goals of the present article. For the purposes of this study, however, I focus on two key aspects of Tungusic pragmatics: word order and, related to it, discourse structure. Nichols finds that word order plays an important role in expressing discourse relations in Tungusic, a claim which is investigated in section 4. Before looking at word order and discourse structure, I first discuss discourse particles, which are not mentioned in her study. At one level this is not surprising: Nichols (1979) focuses on *syntax* and pragmatics; Tungusic discourse particles do not appear to have a syntactic function. Second, because she had to rely on existing description, her analysis could be only as complete as the description itself.

As I show, the existent description and documentation of discourse particles and word order is insufficient. These categories are not directly elicitable, and so I argue for broad documentation of naturally occurring (i.e. spontaneous) conversation and language use in a wide range of registers. Specific questions to be addressed include: how does one guarantee or even work toward a comprehensive documentation of pragmatics? How does one obtain the necessary data for certain language uses (such as discourse particles) which are not directly elicitable? How should they be described in the documentation?

### 3. Discourse particles

Discourse particles have been studied in a number of languages of wider communication<sup>2</sup> but little research has focused on the use of discourse particles in endangered and lesser-studied languages. This is not surprising. Discourse particles are difficult to elicit and hard to describe; moreover, their

---

<sup>2</sup> For a good overview and discussion of the issues involved in defining discourse particles, see Fischer (2000:13-27).

use is rarely mandatory, in the sense that the omission of discourse particles does not normally result in ungrammatical sentences *per se*, although the sentences may be more or less felicitous in a given context. Yet discourse particles are widely used in many languages and their use is one factor in distinguishing fully fluent speakers from less fluent ones. Any truly complete documentation needs to include the use and distribution of discourse particles.

The category of discourse particles is difficult to define and has been the subject of extensive study in languages of wider communication such as English and German, but considerably less so in most endangered languages. For our purposes here, discourse particles are characterised as follows, with the caveat that further research in this area in Tungusic may refine this definition. They are short, typically monosyllabic, and cannot occur as complete utterances. This stems from the fact that they do not carry lexicosemantic meaning and contribute little, if anything, to the propositional content of an utterance. For this same reason, they are omissible: they do not carry denotational value but rather convey the subjective emotional or mental attitude of the speaker to some aspect of the communicative situation. More specifically, this may be an attitude toward the propositional content of the utterance, toward the interlocutor, or toward some other element. Thus the omission of a discourse particle does not result in an ungrammatical sentence, but rather reflects a different subjective evaluation of the propositional content, or changes the context of the utterance in some way. This makes testing their usage complicated. A ‘misuse’ of a discourse particle is more tangibly defined by its use in a pragmatically inappropriate context rather than its omission. For the fieldworker the fact that discourse particles are not directly elicitable is particularly problematic: it often seems to be more a question of random chance as to which particles appear in field sessions and it is difficult to know when one has achieved an exhaustive inventory of the particles. My experience in the field with speakers of both endangered languages and languages of wider communication has been that speakers do not have particularly good intuitions about the use of discourse particles in the abstract, and may even declare that they do not use them, despite evidence to the contrary. Speakers, and semi-speakers in particular, do not always recognise that they use particles, and so often maintain that they have no intuitions about their usage. That said, (fluent) speakers do recognise their misuse, *i.e.*, identify when they are used infelicitously. But again, this “misuse” is generally defined in pragmatic, or meta-pragmatic, terms, and not semantic or syntactic ones.

The ‘meaning’ of discourse particles is sentence-transitional (Abraham 1991) in the sense that their usage presupposes contexts that are not shared by the sentences without those particles. Thus any field situation which attempts to test their usage must be rigorous in defining the contexts in which they are

used. The use of discourse particles is profoundly grounded in a cultural conceptual frame; it is the job of the fieldworker to elicit any and all frames in which a particular particle may be felicitously used. An example of how difficult these are to translate is provided by Arndt (1960:327), which attempts to give the English, German, and Russian equivalents for some particles, as illustrated in (1) with just one particle, German *aber*, Russian *da*, and the English translation:

- |     |         |                                   |
|-----|---------|-----------------------------------|
| (1) | German  | Er weigerte sich <b>aber</b> .    |
|     | Russian | <b>Da</b> on otkazalsja.          |
|     | English | He has refused, <b>mind you</b> . |

In English the meta-pragmatic functions of these particles are usually encoded intonationally rather than morphosyntactically, which adds to the difficulties in translating them directly from a language such as German or Russian. In this particular example, Arndt provided the translation as ‘mind you’, a phrase which in English is marked (both semantically and pragmatically) in a way that the corresponding Russian and German particles are not. The difference in position is also relevant: Russian *da* must be clause-initial; if German *aber* were in first position, it would be better translated as ‘But he refused’; in English the pragmatics of both particles would probably be better conveyed by intonation (and, perhaps, paralinguistic features), inasmuch as English does not have such modal particles, and so any parallel translation that involves a phrase like ‘mind you’ conveys more lexico-semantic meaning than the German or Russian.

Very little research has focused on the use and distribution of discourse particles in the Tungusic languages. Evenki has a large number of discourse particles (Bulatova 1987:710-76; Bulatova and Grenoble 1998:52-53; Konstantinova 1964:251-266). With the exception of *kə*, the particles in Evenki are clitics and occur word-finally but not necessarily at the end of a sentence or clause. They can be used in combination with other particles although the range and limits of such combinations have not been described. The comprehensive list of particles in Evenki includes some which have a conjunctive function and some which have an interrogative function. A number of them have an evaluative kind of function rather than a syntactic one and are identified here as discourse particles. Fluent speakers of Evenki use these particles fairly frequently although not in every utterance; a full documentation of the language clearly requires a description of them and how to use them felicitously. In what follows I briefly discuss two such particles here, *-kV* and *-bo*, in order to illustrate the challenges of adequately documenting them.

Bulatova (1987:71-76) cites 14 of these which are most frequently used. In some cases, she translates them with Russian particles; in others she states that it is difficult to convey the meaning of these particles in Russian. First, the particle *-kV* (*-kal/-kə/-ko*), which Bulatova states corresponds to the Russian particles *a*, *-to* or *že* (Bulatova 1987; Vasilevich 1948:758); part of the problem here is that these Russian “translations” or equivalents are all used differently, and each has multiple pragmatic functions. At the same time, *-kV* is described by Konstantinova (1964:255) as being used for emphasis, or for specifying the meaning of one or another word in the sentence or the entire sentence. (It is unclear exactly how to interpret “emphasis” here, and the examples provided by Konstantinova do not help clarify her description.) Examples are given (2) and (3) from Bulatova (1987:74), along with her translations into Russian, which give the equivalent for Evenki *-kə* in (2) as Russian *že* and for *-kə* in (3) as Russian *a*<sup>3</sup>:

- (2) *i:-lə:-kə*                      *amin-ti*                      *ŋənə-rə-n*  
 where-loc-part                      father-1pl.incl                      go-pst-3sg

R    ‘*Kuda že naš otec pošel?*’  
 ‘Where did our father go?’

- (3) *bi:-kə*                      *nujan-man*                      *o:n*                      *əxim*                      *sara*  
 1sg-part                      3sg-acc                      how                      neg                      know

R    ‘*A ja ego počemu ne znaju?*’  
 ‘How come I don’t know him?’

A comparison of the English and Russian glosses of these examples illustrates some of the inadequacies of the current descriptions. In the Russian of (2), *že* is often described as an emphatic particle, but its use and distribution are quite complicated. (See, for example, discussions of *že* as marking thematic structure in Bonnot 1988; Padučeva 1988; or as signalling assessment of probability in Plungian 1988, among others.) In contrast, Russian *a* (in 3) is a conjunction, usually signalling some kind of contrast. As seen in the English counterparts, the particles simply do not translate into English. As the English

<sup>3</sup>Abbreviations uses in the morpheme-by-morphem glosses are 1: first person; 3: third person; abl: ablative; acc: accusative; all: allative; ant: anteriority; cond: conditional; cvb: converb; dim: diminutive; dat: dative; fut: future; gen: genitive; hab: habitual; incl: inclusive; ipfv: imperfective; loc: locative; neg: negative; p: participle; part: particle; pfv: perfective; pft: perfect; pl: plural; poss: possessive; prs: present; pst: past; Q: interrogative; refl: reflexive; sim: simultaneity ; sg: singular.

glosses indicate, none of these pragmatic subtleties translate adequately into English. English simply does not have modal particles like these; any lexical words used to convey the gist of these particles carry too much lexicosemantic information. Instead, English generally relies on intonation. It is impossible to tell how to use these particles felicitously based on these glosses and/or on the current descriptions.

Part of the problem arises from the fact that the contact language, Russian, is used as a meta-language to describe Evenki. This is, of course, the most usual approach in documentation: the lesser-studied language is documented and described in a language of wider communication. With discourse particles, however, there is a particular danger that the description itself may be unduly influenced by the categories found in the meta-language. For example, the Evenki particle *-bo* is described as not being translatable (Bulatova 1987). Consider its use in example (5), taken from my own field notes (recorded in the village of Iengra, Sakha), here in response to the question posed in the first line of this excerpt:

(5) *mutʃu-ri-tin-ŋu*

return-pst-3pl-part.Q

R: ‘*Oni vernulis*’?’

E: ‘Had they returned?’

*o:n ə-xi-tin mutʃu-tʃa-l bi-xik-tin do:ldi-mtʃa-β-bo*

how neg-prs-3pl return-cond-pl be-cond-3pl hear-cond-1sg-part

*itʃə-mtʃə-l-βun-bo*

see-cond-pl-1pl-part

R: ‘*Kak, net. Esli by vozvratilis*’, *uslišala by, uvideli by*’

E: ‘Oh, no they hadn’t. If they’d returned, I would have heard, we would have seen [them]’

The kind of description which would be useful would comprise an adequate understanding of all circumstances in which it is possible, including full discussion of how the particle does or does not reflect on attitudes of the speaker to the propositional value of the utterance, to the interlocutor(s), to the situation being described, and so on. This kind of information cannot be encapsulated into a gloss, and so would need to be referenced within accompanying materials, i.e. through linked corpora of supporting materials (as suggested in another context by Thieberger 2004)

The point here has not been to provide a full account of these particles but rather to illustrate the difficulties of translating them from one language to another: there is no isomorphic mapping between the pragmatic uses of discourse particles in Evenki and any other language. In fact, their uses appear to differ even in Evenki and closely related Oroqen, although more research is needed on this topic. Fieldwork indicates different inventories of discourse particles in the two, and all Oroqen discourse particles — with the exception of the particle *gə* — occur only sentence-finally.<sup>4</sup> Such differences point to the need for in-depth documentation in order to have a record of these particles and to have some understanding of their use. Because they are not directly elicitable, it is important to collect a wide variety of different registers and genres in different communicative settings. The fact that speakers do not have good intuitions about them and have difficulty articulating their meaning further emphasises the importance of broad documentation. One striking gap in present documentation of all Tungusic languages is conversational data, which is precisely where use of some discourse particles might be anticipated.

At the same time the complexity of discourse particles usage underscores the necessity of some kind of descriptive analysis in basic documentation: without description, the function and use of such particles is opaque. Glossed texts cannot begin to provide the necessary information for second language learners to use particles felicitously, just as they do not provide full information for linguists studying them. Since these particles are in fact not always translatable, it is important to provide an annotated discussion of their uses and the contexts in which they would be appropriate and, optimally, some discussion of when they would be infelicitous. Ultimately a description of discourse particles needs to provide a complete and exhaustive account of the different interpretations and different contexts in which each particle could be used.

This is a prime example of a case where documentation needs to go beyond recording and analyzing discourse in a format which will be accessible to future generations. Instead, it needs to provide some descriptive analysis which takes into account the pragmatics of the sociolinguistic setting.

#### **4. Word order**

The Tungusic languages have Subject Object Verb word order. In Evenki SOV word order is regularly found in folklore and epic texts which were collected in the first half of the 20th century, before extensive language shift had occurred. It is generally found in folklore texts collected today, although

---

<sup>4</sup> Lindsay Whaley, personal communication. The Oroqen particle *gə* (Evenki *kə* or *gə*) ‘well’ is found sentence-initially.

they are usually collected from the oldest generation of speakers who continue to use Evenki as their primary language for communication and see it as their first language. The folklore texts tend to be characterised by the use of formulaic and archaic language; given the lack of textual documentation prior to the beginning of the 20th century, they are the best available indicator of older language patterns. Here the word order is relatively consistently SOV. Similarly, when sentences are elicited in isolation, current speakers consistently produce SOV word order.

Evenki is spoken in an areal context of SOV word order, in contact with other Tungusic languages, as well as contact with Turkic and Mongolic languages, but influence from Russian (an SVO language) has also long been attested. SOV word order is both genetically and typologically consistent for agglutinating languages; given the predominance of SOV languages (i.e. the Altaic languages represented by the Turkic, Mongolic and Tungusic families) in Siberia, it could also be argued that it is areally consistent. If these languages were to exert a strong areal influence, it might be expected that other non-SOV languages spoken in the region would move toward SOV word order. In fact, however, Russian word order has influenced Evenki word order; local Russian dialects show no evidence of a shift toward SOV order. One example of contact-induced word order change in Evenki is found in possessive constructions. Possession is head-marked in Evenki, with the possessive suffix on the possessee NP, as in (6a); in Russian it is dependent-marked as seen in (6b):

(6a) Evenki, head-marked

<i>əβənki-l</i>	<i>oro-r-tin</i>
evenk-pl	reindeer-pl-3pl.poss

(6b) Russian, dependent-marked

<i>olen-i</i>	<i>evenk-ov</i>
deer-pl	evenk-gen.pl

‘The Evenkis’ reindeer’

As (6a) and (6b) illustrate, the prototypical word order in Evenki is possessor-possessee, and in Russian the opposite (possessee-possessor). Under Russian influence, the Evenki word order is sometimes reversed, although morphological marking stays the same, maintaining head-marking:

- (6c) *oro-r-tin*                      *əβəvki-l*  
 reindeer-pl-3pl.poss              evenk-pl  
 ‘The Evenkis’ reindeer’

This change was noted as early as Vasilevich (1948:79) for the Sym dialect, as shown in example (7):

- (7) *tar*                      *dʒur-βa-tin*                      *əβəvki-l*                      *baka-ra-n*  
 3sg.dem                      home-acc-3pl                      evenk-pl                      find-pst-3sg  
 ‘He found the Evenkis’ home’

At that time this word order was exceptional, and is noted by Vasilevich in only this one dialect; (7) provides an unusually clear illustration of Russian influence, inasmuch as that is the only likely explanation for this change. With the gradual loss of Evenki, and the total loss of Evenki monolingualism, the possessee-possessor order has become more widespread and less extraordinary, although the expected order (possessor-possessee) is also commonly found.

Beyond this relatively unambiguous case of contact-induced change, it would appear that Evenki word order can be used for discourse functions such as the marking of topicalization, as discussed for Tungusic as a whole in Nichols (1979). The subject generally occurs sentence-initially but it can follow verb if it is being singled out; often in these cases it is used with *ələ* ‘only’ or with a borrowing from Russian that is similar in meaning (Kolesnikova 1966:179):

- (8) *amaski*]                      *ai ʎt-tʃa-pki*]                      *ələ*                      *saman*  
 formerly                      heal-ipfv-p.hab                      only                      Shaman  
 ‘It used to be only the shaman would heal [people]’

Here the subject *saman* ‘shaman’ follows the verb due to some kind of emphatic stress. In other cases, one would expect SOV word order, but Kolesnikova (1966:177) notes that “Evenki word order is often corrupted under Russian influence; quite frequently Evenki order the sentential components according to Russian norms.” Example (8), the only example of such word order provided by Kolesnikova (1966), is a clear illustration of emphatic word order, with the subject sentence-final. The uses and limits of

such word order at that time are not specified in her description and remain unclear today.

In contrast to Evenki, word order in Oroqen is rigid SOV; in both elicited sentences and in texts, the verb consistently comes at the end of the clause. This is surprising, given the relative flexibility of Evenki word order. It is difficult to determine whether Mandarin word order has had any impact on Oroqen word order. Evenki word order has long been influenced by Russian word order, noted already by Konstantinova (1964). What is unclear is just how much of an impact it has had, and what kinds of constraints there on Evenki word order. The pan-Siberian Tungusic pattern appears to be relatively flexible word order (see Nichols 1979). This is similar to the situation in the primary contact language Russian, in which word order is largely determined by discourse factors such as topicalization and information structure/status. Although ‘neutral’ word order in Russian is SVO, discourse factors often favour a different word order.

Nichols reached this conclusion based on the description for the Tungusic languages available at that time. At present, there would appear to be even less rigid word order in Evenki, again presumably under Russian influence. Even where there is no obvious discourse motivation for changing word order from the SOV pattern, it frequently occurs, as in both (10) and (11) where the dative NP indicating location follows the verb, although typologically and historically it would be expected to precede it:

- (10) *xabal-dʒa-ra-Ø*                      *Jakutskaj-du:*  
 work-ipfv-prs-3pl                      Yakutsk-dat  
 ‘They work in Yakutsk’
- (11) *bi:*            *baldiri:Ø-m*            *Tokorikan-du:*  
 1sg            born-pft-1sg            Tokorikan-dat  
 ‘I was born in Tokorikan.’

Word order here matches what would be expected in Russian, and there is no obvious discourse motivation for the verb not to be in final position in this sentence. Both (10) and (11) illustrate the influence of Russian. In fact, word order changes in contact situations with Russian do not appear to be restricted to the Tungusic languages. In Nivkh (a language isolate spoken in northeast Russia), the canonical word order is rigid SOV but there is a tendency toward free word order in younger speakers, first of all affecting the positions of the subjects and adverbials (Gruzdeva 2000:125). The relatively limited

descriptions in earlier materials on Evenki suggest that the kind of word order illustrated in the examples here is a relatively recent change.

## 5. Discourse structure and tail-head linkage

As a final illustration we will consider the structure of discourse. The overwhelming majority of Tungusic texts collected are folklore or songs, and there has been little analysis of their structure. One discourse device, found with varying consistency in modern Evenki, is what has been called *tail-head linkage*, a term first used by Thurman (1975). The term tail-head linkage is used somewhat differently by different linguists: for some it is used to refer to repetition within a paragraph, while for others that repetition occurs across paragraph boundaries. It is strictly defined by Thompson and Longacre (1985:209) as a device which creates cohesion between successive paragraphs, “something mentioned in the last sentence of the preceding paragraph is referred to by means of back-reference in an adverbial clause in the following paragraph.”

In Evenki and a number of other languages, this kind of repetition tends to be found between sentences but within the body of a chunk of text, i.e. a paragraph, or an analogous topical unit in spoken discourse. In this usage, tail-head linkage consists of a repetition or ‘recapitulation’ of the final clause of the previous sentence or clause chain in the first clause of the next chain or sentence (Genetti 2005; van Kleeef 1988). Usually this involves recapitulation of the final verb of the previous sentence, i.e. the restatement of the verb highlights the main event (van Kleeef 1988:149). In many languages this recapitulation occurs using a different form of the verb in the first (head) clause than that found in the preceding, final (tail) clause. The exact morphology varies from language to language. In Korafe (an SOV Papuan language of Papua New Guinea), the tail-head linkage of lexical verbs across sentence margins signals paragraph-internal sentence junctures. Paragraph junctures are indicated by suspension of tail-head verb recapitulation or by generic verb recapitulation of the ‘tail’ verb” (Farr 1999:337). In Auhelawa (an SOV Austronesian language spoken in Papua New Guinea), a reduplicated form is used in the last verb of a sentence and an unreduplicated form in the next sentence (Lithgow 1995:94-5). The existing descriptions of tail-head linkage, or recapitulation, within a paragraph, suggest that it is a common coherence device in verb-final languages; the position of the verb reinforces the linkage, as it were. More cross-linguistic work is needed to be done in this area.

Tail-head linkage has not been described in the Tungusic languages; in fact, it has not been identified as occurring in them. It is, however, found in Evenki narrative texts, in both traditional folklore and personal narratives. The

final verb in the Evenki “tail” clause is most usually finite, the recapitulated form in the head is often a converb or a participle, although sometimes the tail verb is repeated exactly in the head. Tail-head linkage is widely used in Evenki folklore texts. The following excerpt from the Evenki story *Nimkan* ‘folktale’ (collected in the Sakhalin region; Bulatowa and Cotrozzi 2004:60) is representative:

- (12) 1 *tara:*            *jərək*    *ilara:ka:n*    ***tokorixi-na-n***  
          This        Nerek    3.times-dim    turn-pst-3sg
- 2 ***tokorixin-na-du:n***    *so:*    *gudej*        *du:nna*    *o:-ra-n*  
          turn-p.pft-dat-3sg    very beautiful    land        become-pst-3sg
- 3 *ə-du:*                *jərək*    *dʒolo-tʃi:*        ***garada:-ra-n***  
          here-dat            Nerek    stone-poss        throw-pst-3sg
- 4 ***garada:-na-duki-n***    *so:*    *kətə*    *oro-r*        *xətəkən-tʃə:l*  
          throw-p.perf-abl-3sg    very many    reindeer-pl    jump-p.ant-pl

‘Nerek walked around this [place] three times.

Having walked around it, a very beautiful land appeared.

Here Nerek threw a stone.

Having thrown it, very many reindeer jumped up.’

In (12), the verb of the first sentence *tokorixinan* ‘she walked around’ is recapitulated in the first clause of the next sentence in a participle form (*tokorixinnadu:n*). Similarly, in line 3, the final finite verb (*garada:ran* ‘she threw’) is repeated in the participle form in the beginning of the next clause. In this way the narrative is strung together by the succession of finite past verb form followed by a participle. Such examples are very common in folklore narratives where there is a tight sequencing of events, as in (12). The same usage occurs in conversational narrative texts of some Evenki speakers, as in (13), given by a speaker born in 1930 whose first language is Evenki:

- (13) *bi: buga-la-βi: mutfu-dʒaŋa:-β ənti:l-dulə:-βi:*  
 1sg homeland-loc-refl return-fut-1sg parents-loc-refl  
*dʒu:-la: mutfu:-na əmə-Ø-m dʒan digi-tfi: bi-tfə:-β*  
 home-all return-cvb.sim come-pft-1sg 10 4-poss be-pst-1sg

‘I will return home, to my parents.

‘Returning home, I came, I was 14.’

A review of existing folklore texts and elicited narratives suggests that the use of tail-head linkage in such contexts is relatively predictable in Evenki, thus its absence is striking. The following excerpt comes from an interview with a 12 year old speaker. Despite the sequenced clauses in lines 2-5 of this excerpt, there is no tail-head linkage:

- (14) 1 *ta-du: umuko:n iktə.nə bi-xi-n*  
 there-dat one 3.year.old be-pres-3sg
- 2 *bi: tara dʒaβa-ri:-β*  
 1sg that catch-p.sim.1sg
- 3 *taduk nuŋan-dula:n təg-ri:-m*  
 Then 3sg-loc sit-pst-1sg
- 4 *nuŋan minə-βə garada:-t-ti-n*  
 3sg 1sg-acc throw-pfv-pst-3sg
- 5 *bi: nuŋan-ma:n i:ri:β-dʒa-xa:-i:ri:β-dʒa-xa: tʃutʃuβu-m*  
 1sg 3sg-acc pull-ipfv-cvb.a-pull-ipfv-cvb.a let.go-pst-1sg

1 ‘There is one 3-year-old deer.

2 I caught him.

3 Then I sat on him.

4 He threw me.

5 I pulled and pulled, and then let him go.’

This is the kind of narrative where we would expect to find tail-head linkage: there are tight sequential clauses, and each verb in each of these clauses is telic and signals a completed event. Instead, the complete absence of tail-head linkage is striking. It should be noted that the speaker's first language is Evenki and, at the time of this recording, he was struggling in (the Russian-based) school, in particular in his Russian language classes. In addition, the fluent speakers in the community deemed him to be an excellent speaker, a “**real**” Evenki boy, who speaks “**real**” Evenki. This particular text was recorded in a video-taping session in which three community members participated, all interviewing the boy. After the session they spontaneously and unanimously agreed that he is an excellent speaker. From this we can conclude that either the absence of the use of tail-head linkage was not problematic, or they were exaggerating his proficiency. Assuming the former, then it is unclear if tail-head linkage is something that is acquired as a speaker matures, whether it is simply a question of stylistic preference, or if it is in the process of being lost?

If we turn to Oroqen for comparison, the complete absence of tail-head linkage in all types of discourse, including folklore, is notable. Here again, it is difficult to determine what is the older pattern and what the innovation, nor can we determine if the discourse structure is in any way influenced by the contact languages. In Papua New Guinea, tail-head linkage has been identified as an areal phenomenon, because it occurs across Papuan languages, regardless of typological or genetic differences (Vries 2005). The lack of information about discourse structure in general, and tail-head linkage in particular, across Tungusic languages makes it difficult to determine whether its use in Evenki and lack of use in Oroqen is also areal, the result of contact, attrition, or something else.

## 6. Conclusion

The goal of language documentation is “to provide a comprehensive record of the linguistic practices of a given speech community” (Himmelman 1998:166), and “comprehensive” is intended to include both observable linguistic behaviour and native speakers' metalinguistic and ethnographic knowledge pertaining to language use. The examples provided here—discourse particles, word order variation and discourse structure illustrated by the use of tail-head linkage—underscore the need for comprehensive documentation and description. Moreover, they are indicative of the need to place the socio-cultural setting at the centre of any documentation project and to have the project defined and determined from that centre as its foundation.

## References

- Abraham, Werner 1991. Discourse particles in German: How does their illocutive force come about? In Werner Abraham (ed.) *Discourse particles: Descriptive and theoretical investigations on the logical, syntactic and pragmatic properties of discourse particles in German*, 203-52. Amsterdam: John Benjamins.
- Arndt, Walter 1960. "Modal particles" in Russian and German. *Word* 16:323-36.
- Bonnot, Christine 1988. La particule *že* marqueur de thème. *Les particules énonciative en russe contemporain*, Volume 1, 125-51. Paris: Institut d'études slaves.
- Bulatova, N. Ja. 1987. *Govory èvenkov Amurskoj oblasti*. Leningrad: Nauka.
- Bulatova, N. Ja. and Lenore A. Grenoble 1998. *Evenki* (Languages of the World 141). Munich: Lincom.
- Bulatowa, Nadeshda J. and Stefano Cotrozzi 2004. *Ewenkische Erzählungen aus der Insel Sachalin*. Dresden: Verlag für Kultur und Wissenschaft.
- Burykin, A. A., A. X. Girfanova, A. Ju. Kastrov, Ju. I. Marchenko, N. D. Svetozarova, V. P. Shiff 2005. *Kollekcii narodov Severa v Fonogrammarxive Pushkinskogo doma*. St. Petersburg: St. Petersburg State University.
- Farr, Cynthia J.M. 1999. *The interface between syntax and discourse in Korafe, a Papua language of Papua New Guinea*. Canberra: Pacific Linguistics C-148.
- Fischer, Kerstin 2000. *From cognitive semantics to lexical pragmatics: The functional polysemy of discourse particles*. Berlin: Mouton de Gruyter.
- Genetti, Carol 2005. The participial construction of Dolakhā Newar. Syntactic implications of an Asian converb. *Studies in language* 29:35-87.
- Gorelova, Lililya M. 2002. *Manchu grammar*. Leiden: Brill.
- Gruzdeva, Ekaterina. 2000. Aspects of Russian-Nivkh grammatical interference. In Dickey Gilbers, Jos Schaeken and John Nerbonne (eds.) *Languages in contact*, 121-34. Amsterdam: Rodopi.
- Hill, Jane H. 2006. The ethnography of language and language documentation. In Jost Gippert, Nikolaus P. Himmelmann and Ulrike Mosel (eds.) *Essentials of language documentation*, 113-28. Berlin: Mouton de Gruyter.
- Himmelmann, Nikolaus 1998. Documentary and descriptive linguistics. *Linguistics* 36:161-195.
- Janhunen, Juha 2005. Tungusic: An endangered language family in northeast Asia. *International journal of the sociology of language* 173:37-54.
- Kolesnikova, V. D. 1966. *Sintaksis evenkijskogo jazyka*. Moscow-Leningrad: Nauka.
- Konstantinova, O. A. 1964. *Evenkijskij jazyk*. Moscow-Leningrad: AN SSSR.
- Lithgow, David R. 1995. Reduplication for past actions in Auhelawa. *Language and linguistics in Melanesia* 26:89-95.

- Mey, Jacob 1993. *Pragmatics*. Oxford: Basil Blackwell.
- Nichols, Johanna 1979. Syntax and pragmatics in Manchu-Tungus languages. *Chicago Linguistic Society Parasession on the Elements*, 420-28.
- Padučeva, E. V. 1998. La particule *že*: sémantique, syntaxe, prosodie. *Les particules énonciative en russe contemporain*, Volume 3, 11-44. Paris: Institut d'études slaves.
- Plungian, V. A. 1988. Signification de la particule *že* et jugement de probabilité. *Les particules énonciative en russe contemporain*, Volume 3, 45-58. Paris: Institut d'études slaves.
- Silverstein, Michael 2004. "Cultural" concepts and the language-culture nexus. *Current anthropology* 45:621-52.
- Thieberger, Nicholas 2004. Documentation in practice: developing a linked media corpus of South Efate. In Peter K. Austin (ed.) *Language documentation and description*, Volume 2, 169-78. London: School of Oriental and African Studies.
- Thompson, Sandra A. and Robert E. Longacre 1985. Adverbial clauses. In Timothy Shopen (ed.) *Language typology and syntactic descriptions. Volume II: Complex constructions*, 171-234. Cambridge: Cambridge University Press.
- Thurman, R. C. 1975. Chuave medial verbs. *Anthropological Linguistics* 17:342-52.
- van Kleeef, Sjaak 1988. Tail-head linkage in Siroi. *Language and linguistics in Melanesia* 20:147-56.
- Vasilevich, G. 1948. *Očerki dialektov èvenkijskogo jazyka*. Moscow-Leningrad: Nauka.
- Vries, Lourens de 2005. Towards a typology of tail-head linkage in Papuan languages. *Studies in language* 29:363-84.
- Whaley, Lindsay J. 2003. The future of indigenous languages. *Futures* 325: 961-73.
- Whaley, Lindsay J., Lenore A. Grenoble and Fengxiang Li 1999. Revisiting Tungusic classification from the bottom up: A comparison of Evenki and Oroqen. *Language* 75:286-321.