

Northern Khanty demonstratives: a markedness-based approach

Lera Marinina
Individual researcher
vlrmarinina@gmail.com

Abstract

This paper presents a semantic analysis of the Northern Khanty demonstratives. Northern Khanty has a tripartite demonstrative system: proximal term *tām*, distal term *tum* and unmarked one *ši*. All terms may be used both exophorically and anaphorically. Based mainly on Ahn's (2022) and Reisinger and Huijsmans' (2021) approaches to the demonstratives, I account for the data postulating allosemy and presuppositions of different strengths for Northern Khanty demonstratives.

1 Introduction

1.1 Language and speakers

The current paper provides descriptive data and analysis of demonstratives in Kazym dialect of Northern Khanty language (Khanty < Uralic; Russia). According to Kaksin (2010) Kazym dialect is spoken in the upper reaches of the Ob river in Russia by about 1700 people. The data were collected during fieldwork in the village of Kazym (Khanty-Mansi Autonomous Region - Yugra, Russia) in 2022-2024.

1.2 Methodology

All the consultants I worked with are bilingual, i.e., they speak Northern Khanty and Russian, and their first acquired language is Northern Khanty. Russian was used as lingua franca during fieldwork. I use several methods in my research:

- elicitation
- for exophoric uses: elicitation from Russian-language stimuli based on Wilkins' (1999b) experimental questionnaire (distinctions in anchoring, distance, and visibility)
- for contrastive exophoric uses: elicitation from Russian-language stimuli based on experimental questionnaires of Rostovtsev-Popel (2009) and Wilkins (1999a)
- analysis of corpus examples
- quasi-experimental task (referential communication: instructing another person to arrange the objects in the proper order)

2 Northern Khanty demonstrative system

2.1 An overview

Northern Khanty is an articleless language. It has 3 demonstrative bases. The discussion in this paper is mainly concerned with the adnominal forms *tām* (proximal), *tum* (distal) and *ši* (unmarked). There exist a special pronominal form for the unmarked term *šit*, and some speakers also have special pronominal forms *tāmi* and *tumi*. No differences between adnominal

and pronominal forms were observed, besides syntactic differences. All three demonstratives may be used exophorically and anaphorically; no other uses were detected. Table 1 presents basic distinctions that Northern Khanty demonstratives reveal.

Table 1: The inventory of Northern Khanty demonstratives

Demonstrative	Deictic center	Spatial meaning
<i>tām</i>	Speaker	Proximal
<i>tum</i>	Speaker&Addressee	Distal
<i>śi</i>	Speaker&Addressee	Unmarked

I also want to notice that unmarked term *śi* has grammaticalized into the emphatic particle (for instance, see ex. (10) below). I consider it to be a distinct lexeme synchronically and do not discuss it in this paper.

2.2 Exophoric uses

All three demonstratives in Northern Khanty may be used exophorically. Practically, speakers tend to volunteer *tām* and *tum* wherever possible, and *śi* is more often accepted by the consultants than volunteered¹. Proximal demonstrative *tām* is usually used for the referents that speaker may reach with the arm, cf. examples (1) and (2) with reachable and unreachable referent respectively:

- (1) *tām*/#*śi*/#*tum* kinškaj-en ma λəŋət-s-əm.
 this/#DEM/#that book-POSS.2SG I read-PST-1SG
 ‘I read this book.’ (a book in front of the speaker, context 6 from Wilkins (1999b), further cited as WQ6)²
- (2) #*tām*/#*śi*/^{OK}*tum* kinškaj-en ma λəŋət-s-əm.
 #this/#DEM/^{OK}that book-POSS.2SG I read-PST-1SG
 ‘I read this book.’ (WQ12: speaker and addressee next to each other, a book is to the side of the addressee and is unreachable for the speaker)

However, the scale of proximity is relative and may be expanded dependent on referent type and positions of the interlocutors. For example, *tām* is used referring to large referents that are close enough but inaccessible with an arm (e.g. a house several meters away) and to small referents that are between speaker and addressee but unreachable for both of them.

Distal demonstrative *tum* is in complementary distribution with proximal term *tām*. *Tum* is used for the referents outside the shared space³ (see (2) above and (3) below):

- (3) muj pāta #*tām*/^{OK}*śi*/^{OK}*tum* kinškaj-en kamən kərət’λ’ə-λ?
 why #this/^{OK}DEM/^{OK}that book-POSS.2SG outside lie-NPST[3SG]
 ‘Why is that book lying outside?’ (WQ21: a book is several meters away from speaker and addressee)

Unmarked demonstrative *śi* is not in complementary distribution with *tām* and *tum*. In fact, it is allowed in all contexts where distal *tum* is used (e.g. (2) and (3) above). Regarding the

¹In the examples I use no symbol for the volunteered variant and OK for the accepted one.

²In this example and in the majority of the following ones: the possessive marker of 2sg -en functions as a salient article, see Mikhailov (2024) for the discussion.

³Shared space is the space in-between the interlocutors.

variation of *tām* and *śi*, *śi* is alternative to *tām* in the situations when the referent is reachable for both interlocutors, but is located closer to the addressee:

- (4) OK_{tām}/OK_{śi}/#_{tum} kinškaj-en nəŋ-en?
 OK_{this}/OK_{DEM}/#_{that} book-POSS.2SG you-DAT
 ‘Is this book yours?’ (WQ9: speaker and addressee sitting next to each other, a book is in front of the addressee)

When proximal and distal terms are unsuitable, *śi* arises as the only one possible option:

- (5) #_{tām}/#_{śi}/#_{tum} kinškaj-en χυλśa wɛ-s-ən?
 #_{this}/DEM/#_{that} book-POSS.2SG from.where take-PST-2SG
 ‘Where did you take that book from?’ (WQ16: several meters between speaker and addressee, the referent is near the addressee)

In contrastive uses with two referents *tām* and *tum* are volunteered regardless of the addressee’s position; *śi* is rarely volunteered, but possible (ex. (6) below). This fact also speaks in favor of unmarkedness of *śi* (see Meira and Terrill (2005) for the discussion of unmarked demonstratives in contrastive contexts).

- (6) *tām*/#*śi*/#*tum* an-ən pun-man šaj, #*tām*/OK_{śi}/*tum* an-ən pun-man jin̄k.
 this/#DEM/#that cup-LOC put-CVB tea #this/OK_{DEM}/that cup-LOC put-cvb water
 ‘This cup has tea in it, and that cup has water in it.’ (Context based on Rostovtsev-Popel (2009), the first cup is closer to the speaker than the second cup)

Summing up, proximal *tām* and distal *tum* encode clear distance distinctions, while *śi* may be used in a wide range of contexts and is semantically vague with respect to the distance.

2.3 Attention and salience

According to Diessel (2006), there are two main functions of demonstratives: indicating the location of a referent and coordinating the focus of interlocutors’ joint attention. Since *śi* is unmarked in distance, I hypothesized that its functions are related to attention, i.e. it is favored for the referents in the joint focus of attention. This hypothesis was confirmed only partly. In quasi-experiment, *śi* was not used for drawing addressee’s attention to the referent at all. Instead, it was often used in the contexts of confirmation like (7), with the referents in the focus of joint attention:

- (7) [There are two sweets that are not arranged yet. Participant looks at the photo and then points to the position near the previously placed object on the toy table]
 Participant:
 maw tiw pun-e.
 sweet here.ILL put-IMP.SG>SG
 ‘Put [some] sweet here.’
 Assistant: [looks at two sweets that are not arranged yet and takes one of them]
 Participant (confirming):
 śi maw-en.
 DEM sweet-POSS.2SG
 ‘This sweet [is the one I intended].’

However, in other (non-confirming) contexts with referents in joint attention *śi* is not preferable: the choice of the demonstrative rather depends on distance to the referent, as in (8). It seems that in examples like (7) anaphoricity and not attentional focus matters (see the next subsection

for the discussion of anaphoric uses).

- (8) [You visit me for the first time, and you do not know that I have a dog. We are sitting in a room and talking. Suddenly a dog runs into the room and starts barking. We look at it, then I say:]
 aλ pǎλ-a, tǎm/śit/tum ma amp-εm.
 PROH be.afraid-IMP[SG] this/DEM/that I dog-POSS.1SG
 ‘Don’t be afraid, this/that is my dog.’

Thus, focus of attention itself turns out not to be a distinctive parameter for Northern Khanty demonstrative system. I discuss in Section 3.1 why in spontaneous speech *śi* is not chosen for drawing attention.

2.4 Anaphoric uses

Unmarked term *śi* is natural in anaphoric uses (in the sense of Ahn (2022)) (ex. (9) below), while *tǎm* and *tum* rather require specific conditions. Anaphoric *tǎm* and *tum* arise in the contexts with more than one salient referent as (10): *tǎm* is used for a more salient referent (main character of the story), and *tum* for a less salient one (a secondary character).

- (9) [At the table people are discussing a woman who lives in Novosibirsk and studies Khanty. You do not know her, but you heard that she came to Kazym a few years ago. You say:]
 śi/#tǎm/#tum imij-en Amña woš-a juχət-s.
 DEM#this/#that woman-POSS.2SG Amnya town-DAT come-PST[3SG]
 ‘That woman visited Kazym.’
- (10) [The speaker tells a folklore story. The main character could not sleep that night and saw a creature. It touched another man who woke up frightened.]
 in tǎm ike-λ pa śi nuχ oms-əmt-əs inśəs-λ-əλλe: “χuti
 now this man-POSS.3SG ADD EMPH up sit-PNCT-PST[3SG] ask-NPST-3SG>SG how
 ji-s-ən?” in tum ik-en lup-ijəλ-λ: “pa muλsər
 become-PST-LOC now that man-POSS.2SG say-FREQ-NPST[3SG] ADD what.INDEF
 pun-əŋ još-əp ut-ən pəχəλ-mə-s-ij-əm.”
 hair-PROP hand-PROP.P thing-LOC poke-MOM-PST-PASS-1SG
 ‘And this man [who told a story] sat and asked him: “What happened?” And that man [who woke up frightened] says: “Someone with a hairy hand poked me.”’

3 Analysis

3.1 Current analysis

I combine features of the analyses of Reisinger and Huijsmans (2021) and Ahn (2022). Ahn argues that demonstratives introduce an additional syntactic argument R that can be occupied by a pointing gesture, an anaphoric index or a relative clause. In Northern Khanty R may be occupied by pointing gesture and anaphoric index only. I assume that this happens due to syntactic restrictions and do not discuss it further here. Following Reisinger&Huijsmans’ analysis, I include spatial and salience components into the presuppositions of demonstratives. In the analysis I want to capture several crucial facts. Firstly, all three demonstratives may be used exophorically, *tǎm* and *tum* are preferred in contrastive contexts. Secondly, *śi* is rarely volunteered in the contexts of drawing the addressee’s attention to the referent. Thirdly, *śi* is preferred in anaphoric uses, and *tǎm* and *tum* can be used anaphorically if there is more than one salient referent.

I propose to analyze *śi* as an unmarked demonstrative and postulate an underspecified presupposition for it. The referent must meet the descriptive content of N, and the R slot must be occupied, while no spatial or salience information is included:

$$(11) \quad \llbracket \acute{s}i \rrbracket^c = \lambda N. \lambda R. \iota y. N(y) \wedge R(y)$$

For *tăm* and *tum* I propose additional presuppositions of proximity and distality respectively (defined below):

$$(12) \quad \llbracket t\grave{a}m \rrbracket^c = \lambda N. \lambda R. \iota y. N(y) \wedge R(y) \wedge PROX(y)$$

$$(13) \quad \llbracket tum \rrbracket^c = \lambda N. \lambda R. \iota y. N(y) \wedge R(y) \wedge DIST(y)$$

I propose that PROX and DIST are subject to allosemy: an appropriate interpretation for PROX/DIST is chosen during Spell-Out of the syntactic structure to Logical Form depending on the syntactic context of the demonstrative head. In our case, the conditioning factor is whether GestP, i.e., a gesture phrase, or IdxP, an index phrase, is in the R position:

- (14) a. PROX = $\lambda x. \lambda N. x$ is close in space to l_c / _____ GestP
 b. PROX = $\lambda x. \lambda N. x \in Sal_c \wedge \exists y [y \in Sal_c \wedge N(y) \wedge y <_{sal} x] \wedge \forall z [z \in Sal_c \wedge N(z) \wedge z \neq x \rightarrow z <_{sal} x]$ / _____ IdxP
- (15) a. DIST = $\lambda x. \lambda N. x$ is far in space from l_c / _____ GestP
 b. DIST = $\lambda x. \lambda N. x \in Sal_c \wedge \exists y [y \in Sal_c \wedge N(y) \wedge x <_{sal} y]$ / _____ IdxP

I use a contextual parameter l_c — the location of the utterance situation. I do not provide a more explicit definition for ‘close’ and ‘far’ since these concepts are to a large extent vague, as the data on exophoric uses show.

Informally, (14) means that if there is GestP in the R position, for *tăm* to be felicitous the referent x (denoted by *tăm*) must be located close to the location of the utterance situation l_c ; if there is IdxP in the R position, for *tăm* to be felicitous the referent x must be the most salient in the context c (all other salient referents that meet the descriptive content of N are less salient), and there must be at least one referent that is less salient than x . Similarly, (15) means that if there is GestP in the R position, for *tum* to be felicitous the referent x (denoted by *tum*) must be located far from l_c ; if there is IdxP in the R position, for *tum* to be felicitous the referent x (denoted by *tum*) must be salient in the context c , and there must be at least one referent that is more salient than x .

We can now appeal to *Maximize presupposition!* (Heim 1991) to explain the distribution in (1)-(6): if there are several contextually equivalent alternative elements, the element with the stronger presuppositions must be chosen. Applying *Maximize presupposition!* allows us to account for the fact that exophoric *śi* is rarely volunteered in the contexts of drawing attention (see Section 2.3). At the same time, wide distribution of *śi* in both exophoric and anaphoric contexts is predicted: a weaker presupposition of *śi* is less likely to fail, compared with presuppositions of *tăm* and *tum*.

In exophoric uses, *tăm* is predicted to be used for proximal referents and *tum* for distal ones, while *śi* must be preferred for referents that are neither really close to l_c nor quite far enough. In anaphoric uses, if there is more than one salient referent meeting the descriptive content of N, according to *Maximize presupposition!* *tăm* will be used for the most salient referent, and *tum* for the less salient one (ex. (10)). But if there is only one salient referent, the presuppositions of *tăm* and *tum* fail, so *śi* will be used instead (ex. (9)). Another correct prediction that the analysis makes is that *śi* is preferred for confirmation (ex. (7)): such contexts are treated as anaphoric with one salient referent.

3.2 Other predictions and challenges for the analysis

The proposed analysis with minor speculations lets us account for temporal deixis. Temporal uses of demonstratives may be analyzed as anaphoric. If the referent is a current time interval, it is the most salient referent among all salient referents of certain type, so the presupposition of anaphoric *tām* is satisfied, as in (16). If the referent is some other time interval, it is not clear if it is more or less salient than the current time interval, so presuppositions of *tām* and *tum* fail, and *śi* is used, as in (17).

- (16) *tām/#śi/#tum jetn wən wot-a ji-λ.*
 this/#DEM/#that evening strong wind-DAT become-NPST[3SG]
 ‘There will be a strong wind this evening.’ (The referent is the most salient evening of all the evenings.)
- (17) [Starting the story]
śi/#tām/#tum jetn wotas wə-s.
 DEM/#this/#that evening blizzard be-PST[3SG]
 ‘There was a blizzard that evening.’ (The evening under discussion is equally salient to current evening.)

However, there are some problems for the analysis. Firstly, it does not capture variability, as applying *Maximize presupposition!* implies having only one option in certain context. We may assume that variability is a problem of elicitation or that the speakers accommodate presuppositions, but these explanations are not always satisfactory. Secondly, the analysis does not capture for some contrasts detected in the new field data.

tām is preferred in anaphoric uses when the referent is present in utterance situation and is located close to the speaker:

- (18) [I am walking through Kazym with a new bicycle when I meet a friend. I say:]
 want-e majsər welosipet təj-λ-əm! *tām/%%śi/#tum*
 look-IMP.SG>SG what bicycle have-NPST-1SG this/%%DEM/#that
 welosiped-əm aŋk-əm-a mojlə-λ-əm.
 bicycle-POSS.1SG mother-POSS.1SG-DAT gift-NPST-1SG>SG
 ‘Look, what a bicycle I have! I will give this bicycle as a present to my mother’

Within our analysis, could we treat such uses as exophoric if *tum* behaved the same way. However, *tum* is prohibited in anaphoric contexts when the referent is present in the discourse situation and is far from the speaker; in such contexts *śi* is used. Additionally, there is some evidence that distal *tum* used exophorically requires visual landmarks or another salient referent in the context. Thus, further investigation is needed and the analysis presented here needs to be modified, although the main ideas are likely to remain the same.

4 Conclusions

The proposed analysis unifies two crucial uses of demonstratives, exophoric and anaphoric. It accounts for the main contrasts of the demonstrative system in Northern Khanty and proposes a new idea of limiting some demonstratives to contexts of salience-driven competition. However, current version of the analysis does not catch some data, so it needs modifications. It also does not focus on distance distinctions of exophoric demonstratives; a distinct analysis in the framework of Optimality Theory might be more applicable for the purposes of predicting the distribution in exophoric uses.

References

- Ahn, Dorothy (2022). “Indirectly direct: An account of demonstratives and pointing”. In: *Linguistics and Philosophy* 45, pp. 1345–1393.
- Diessel, Holger (2006). “Demonstratives, joint attention, and the emergence of grammar”. In: *Cognitive Linguistics* 17.4, pp. 463–489.
- Heim, Irene (1991). “Artikel und Definitheit”. In: vol. 17. De Gruyter Mouton.
- Kaksin, Andrey (2010). *Kazymskij dialekt khantyjskogo yazyka. (Russian) [Kazym dialect of Khanty language]*. Yugorskii gosudarstvennyi universitet.
- Meira, Sérgio and Angela Terrill (2005). “Contrasting contrastive demonstratives in Tiriyo and Lavukaleve”. In: *Linguistics* 43.6, pp. 1131–1152.
- Mikhailov, Stepan (2024). “Semantics of the Northern Khanty Salient Article: Definiteness, salience, and obviation”. In: *Voprosy jazykoznanija* 1, pp. 7–38.
- Reisinger, Daniel and Marianne Huijsmans (2021). “Demonstratives in ʔayʔajuθəm”. In: UBCWPL.
- Rostovtsev-Popel (2009). “Tipologiya demonstrativov: srednie dejktiki”. In: *Voprosy jazykoznanija* 2, pp. 22–34.
- Wilkins, David (1999a). “Eliciting contrastive use of demonstratives for objects within close personal space (all objects well within arm’s reach)”. In: Max Planck Institute for Psycholinguistics.
- (1999b). “The 1999 demonstrative questionnaire: “This” and “that” in comparative perspective”. In: Max Planck Institute for Psycholinguistics.