

Reference Tracking Mechanisms and automatic annotation based on Ob-Ugric Information Structure

Stefanie Eckmann and Gwen Eva Janda

Ludwig-Maximilian-University of Munich

gwen.janda@lmu.de

December 2nd 2016

Overview

- 1 Ob-Ugric Languages
- 2 Linguistic Typology
- 3 Information Structure
- 4 Reference Tracking
- 5 Reference Tracking Mechanisms
Passive
Dative Shift
- 6 Automatic Annotation

Reference
Tracking and
Information
Structure in
Ob-Ugric

Stefanie
Eckmann
and Gwen
Eva Janda

Ob-Ugric
Languages

Linguistic
Typology

Information
Structure

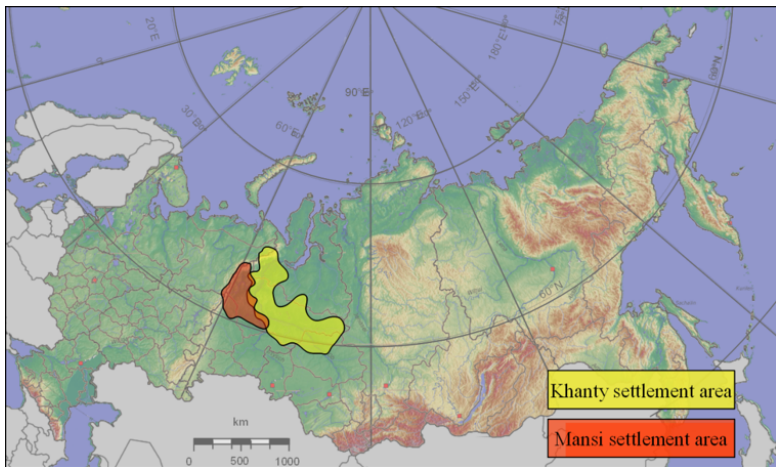
Reference
Tracking

Reference
Tracking
Mechanisms
Passive
Dative Shift

Automatic
Annotation

Ob-Ugric Languages

- branch of the Uralic language family
 - Khanty and Mansi (dialects)
 - spoken in Western Siberia, Ob + tributaries
 - highly endangered languages



Linguistic Typology

- mainly agglutinative
 - subject + object agreement on the verb
 - person as nominal category
 - large inventory of adverbial cases and/or postpositions
- marking of syntactic core roles by
 - word order (SOV)
 - verbal inflection
- tendency of leaving subject and object role unmarked
- pro-drop languages

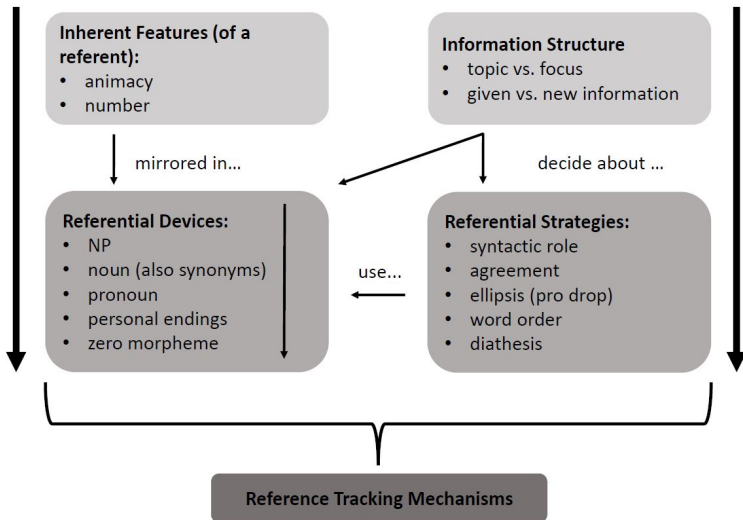
Information Structure

- discourse = grammar + communicative value
- communicative value is based on
 - text coherence and
 - balanced content of given and new information
- ⇒ information structure
- information structure is determined by
 - speaker's intention and
 - speaker's assumption of the hearer's knowledge

Reference Tracking

- monitoring a participant in an on-going discourse (Nagaya 2006: 3)
 - using the principle of coreference
- use of referential devices and strategies
 - based on information structure
 - including inherent features of a referent
- the referents to be tracked are topical
- reference tracking = visualization of information structure in a text

Reference Tracking



Reference Tracking Mechanisms

- several kinds of topics depending on their involvement in the discourse:
 - primary topic (TOP1) (Nikolaeva 2001)
 - secondary topic (TOP2) (Nikolaeva 2001)
 - sentence topic(s) (Reinhard 1982)
- reference tracking mechanisms monitor TOP1 as well as TOP2 referents

Reference Tracking Mechanisms

- strong correlation between syntactic and pragmatic role

syntactic hierarchy:	S	DO	IO	ADV
pragmatic hierarchy:	Top1	Top2	...	

- overrules the correlation between syntactic and semantic role

semantic hierarchy:	AG	PAT	REC	...
---------------------	----	-----	-----	-----

→ diathesis changes

TOP1: Passive

- subject = patient
- referent in agent role is not mandatory
 - syntactic non-core role if mentioned
 - Mansi: DLAT
 - Khanty: LOC

(1) OUIDB Pelym Mansi Corpus. Text ID 1277, Nr. 27

oꝝmpnə purx itwəs.

∅	oꝝmp -nə	pur -x	it -w -əs -∅
∅	dog -DLAT	bite -INF	want -PASS -PST -3SG
S/PAT	ADV/AG		

A dog wanted to bite him.

TOP1: Passive

- subject = recipient/beneficient

(2) OUIDB Surgut Khanty Corpus. Text ID 734, Nr. 20
tot tʌβətət məʎo.

∅	tot	tʌβət -ət	mə ʎ-o
∅	there	fire -INSC	give -PRS -PASS.2SG
S/REC		ADV/PAT	

There you will be given with fire.

TOP1: Passive

- subject = locative (Kulonen 1989)
- recipient in sentences with verbs of motion

(3) OUIDB Pelym Mansi Corpus. Text ID 1335, Nr. 11
ta ke:m wuləmnə tʲi joxto:ws.

∅	ta ke:m	wuləmnə	joxto -w -s -∅
∅	to an extent	sleep -DLAT	come -PASS-PST -3SG
S/LOC		ADV/AG	

She was completely overcome by sleep.

TOP1: Passive

- subject = ADR
- verba dicendi

(4) OUIDB Surgut Khanty Corpus. Text ID 734, Nr. 20
mʌβəʈinə təyə mu:nʲtʲo.

∅	mʌβəʈi -nə	təyə	mu:nʲtʲ -∅ -o
∅	what -LOC	here	tell tales -PST -PASS.2SG
S/ADR	ADV/AG		

What has told you here?

TOP1: Passive

- Any semantic role can be promoted in subject position
 - patient is not promoted
 - agent is not demoted
 - but TOP1 remains subject
- transitivity of the verb is not decisive for passivization
- passive is a reference-tracking mechanism used regardless of verb valency and semantic roles

TOP2: Dative Shift

- ditransitive sentence

(5) OUIDB Surgut Khanty Corpus. Text ID 1083, Nr. 22

n#ŋ mə:ntem məje ɛ:j nʲe:βreməle.

n#ŋ	mə:ntem	məj -e	nʲe:βrem -əle
2SG	1SG.DAT	give -IMP.SG>2SG	child -DIM.MEL
S/AG	IO/REC		DO/PAT

You give me your [small] child!

TOP2: Dative Shift

- Dative Shift

(6) OUIDB Surgut Khanty Corpus. Text ID 1083, Nr. 22

tʲu:t me: nʌŋet tʌβətet mətəm.

tʲu:t	me:	nʌŋet	tʌβət -et	mə -t-əm
then	1SG	2SG.ACC	fire -INSC	give -PRS -1SG
	S/AG	DO/REC	ADV/PAT	

Then I will give you fire.

TOP2: Dative Shift

- marking of dative shift in SK:
 - DO/Recipient with ACC (pronominal)
 - ADV/Patient with INSC
- marking of dative shift in NM:
 - DO/Recipient unmarked (pro-drop)
 - ADV/Patient with INST
 - objective conjugation

TOP2: Dative Shift

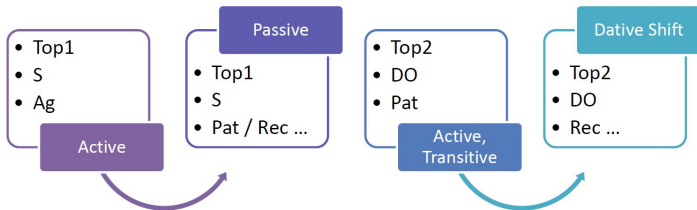
- (7) OUIDB Northern Mansi Corpus. Text ID 1229, Nr. 6
n^ja:ləl wa:rɪlum

∅	∅	n ^j a:l -əl	wa:r -i -lum
∅	∅	arrow -INST	make -PRS -SG<1SG
S/AG	DO/REC	ADV/PAT	

I make you an arrow.

- dative shift is a reference tracking mechanism to promote TOP2 into DO position

Summary



Automatic Annotation

- Ob-Ugric sentence structure exhibits regularities based on information structure
- these regularities can be used to develop (semi-)automatic annotation tools for texts
- DFG/FWF project Ob-Ugric Database (OUIDB) – www.ouidb.gwi.uni-muenchen.de
 - functional
 - semantic
 - pragmatic
 - referential annotation

Basis for Automatic Annotation

Reference
Tracking and
Information
Structure in
Ob-Ugric

Stefanie
Eckmann
and Gwen
Eva Janda

Ob-Ugric
Languages

Linguistic
Typology

Information
Structure

Reference
Tracking

Reference
Tracking
Mechanisms

Passive
Dative Shift

Automatic
Annotation

- correlation of syntactic and pragmatic role
- agreement on the verb:
 - subject role
 - direct object role (objective conjugation)
- ellipsis (pro-drop):
 - topics in subject position can be dropped (TOP1)
 - TOP in direct object position triggers object agreement on the verb
 - → TOP2 can be dropped

Annotation Rules

- if the sentence consists only of a verb
→ zero anaphora occurs
- the sentence consists only of a verb with object-verb agreement
→ two zero anaphora occur

∅

∅

n̥a:ləl

wa:rilum

n̥a:l-əl

wa:r-i-lum

arrow-INST

make-PRS-SG<1SG

subs-infl:n

v-infl:v-infl:v

[33]

[34]

zero

zero

NP

okVP

Annotation Rules

- word order: zero anaphora occurs sentence-initial
 - first zero in the sentence is tagged S AG and TOP
 - second zero in sentence is tagged DO PAT and TOP

∅

∅

n̩a:ləl

wa:rillum

n̩a:l-əl

wa:r-i-lum

arrow-INST

make-PRS-SG<1SG

subs-infl:n

v-infl:v-infl:v

[33]

[34]

zero

zero

NP

okVP

S	func
---	------

O	func
---	------

ADV	func
-----	------

PRED	func
------	------

AG	sem
----	-----

REC	sem
-----	-----

PAT	sem
-----	-----

	sem
--	-----

TOP	pra
-----	-----

TOP	pra
-----	-----

FOC	pra
-----	-----

FOC	pra
-----	-----

Reference
Tracking and
Information
Structure in
Ob-Ugric

Stefanie
Eckmann
and Gwen
Eva Janda

Ob-Ugric
Languages

Linguistic
Typology

Information
Structure

Reference
Tracking

Reference
Tracking
Mechanisms

Passive
Dative Shift

Automatic
Annotation

Thank you for your attention

References

- Givón, Talmy. 1984. "Direct Object and Dative Shifting: Semantic and Pragmatic Case". *Objects. Towards a theory of grammatical relations*. London: Academic Press. p. 151-182.
- Kibrik, Andrej. 2008. "Reference maintenance in discourse". *Language Typology and Language Universals*. p. 1123-1141.
- Kulonen, Ulla-Maija. 1989. *The Passive in Ob-Ugrian*. In: MSFOu 203. Helsinki: Suomalais-Ugrilainen Seura.
- Kulonen, Ulla. 1986. "On the functions of the passive suffix *-we in Vogul and Finnish". *FUF* 47. Helsinki: Suomalais-Ugrilainen Seura. p. 223-230.
- Nagaya, Naonori. 2006. "Topicality and Reference-Tracking in Tagalog". *Presented at the 9th Philippine Linguistics Congress, University of the Philippines Diliman*.
- Nikolaeva, Irina. 2001. "Secondary topic as a relation in information structure". *Linguistics* 39,1. p. 1-49.
- Reinhart, Tanya. 1982. "Pragmatics and linguistics: An analysis of sentence topics". *Philosophica* 27. p. 53-94.
- Skribnik, Jelena. 2001. "Pragmatic Structuring in Northern Mansi". *Congressus IX Fenno-Ugristarum*. p. 222-239.